

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2004, 01:33:43 ; Search time 45 Seconds  
(without alignments)  
2455.024 Million cell updates/sec

Title: US-09-854-122-16  
Perfect score: 2059  
Sequence: 1 TRGITGIAVACLEPLIMAGIL.....CLLSYFVINPILLKKKK 391

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq\_29Jan04:.\*  
1: Geneseqp1980s:.\*  
2: Geneseqp1990s:.\*  
3: Geneseqp2000s:.\*  
4: Geneseqp2001s:.\*  
5: Geneseqp2002s:.\*  
6: Geneseqp2003as:.\*  
7: Geneseqp2003bs:.\*  
8: Geneseqp2004s:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1996	96.9	397	5 AAU12056	Aau12056 Protein c
2	1750	85.0	331	5 AAU12057	Aau12057 Zostera m
3	572.5	27.8	331	3 AAG21657	Aag21657 Arabidops
4	572.5	27.8	341	3 AAG21656	Aag21656 Arabidops
5	567	27.5	346	5 ABB91056	Abb91056 Arabidops
6	567	27.5	346	6 ABB91056	Abb91056 Arabidops
7	562.5	27.3	338	6 ABB91056	Abb91056 Arabidops
8	561.5	27.3	338	3 AAG16972	Aag16972 Arabidops
9	558.5	27.1	350	3 AAG17409	Aag17409 Arabidops
10	558.5	27.1	350	5 ABB91589	Abb91589 Arabidops
11	558.5	27.1	350	6 ABB91589	Abb91589 Arabidops
12	555.5	27.0	359	3 AAG46836	Aag46836 Arabidops
13	555.5	27.0	359	3 AAG44205	Aag44205 Arabidops
14	555.5	27.0	359	4 AAB49722	Aab49722 Hydroxyja
15	551.5	26.8	354	3 AAG44206	Aag44206 Arabidops
16	551.5	26.8	354	3 AAG46837	Aag46837 Arabidops
17	551.5	26.8	354	5 ABB93382	Abb93382 Arabidops
18	550.5	26.7	327	3 AAG16879	Aag16879 Arabidops
19	550.5	26.7	337	3 AAG16878	Aag16878 Arabidops
20	550.5	26.7	349	3 AAG46838	Aag46838 Arabidops
21	550.5	26.7	349	3 AAG44207	Aag44207 Arabidops
22	548.5	26.6	351	3 AAG17285	Aag17285 Arabidops
23	545	26.5	318	3 AAG17287	Aag17287 Arabidops
24	545	26.5	331	3 AAG17286	Aag17286 Arabidops
25	545	26.5	331	5 ABB91710	Abb91710 Arabidops

26	544.5	26.4	326	5 ABB91155	Abb91155 Arabidops
27	537.5	26.1	326	3 AAG29670	Aag29670 Arabidops
28	537.5	26.1	326	5 ABB91711	Abb91711 Arabidops
29	537.5	26.1	326	6 ABB91711	Abb91711 Arabidops
30	536.5	26.1	351	5 ABB90981	Abb90981 Arabidops
31	525.5	25.5	333	5 ABB91753	Abb91753 Arabidops
32	510.5	24.8	340	5 ABB93381	Abb93381 Arabidops
33	510.5	24.8	347	4 AAB49723	Aab49723 Hydroxyja
34	486.5	23.6	314	5 ABB93110	Abb93110 Arabidops
35	483	23.5	688	7 ADD25205	Add25205 Fertilty
36	407	19.8	273	5 ABB91902	Abb91902 Arabidops
37	393.5	19.1	258	6 AAE35941	Aae35941 Lolium pe
38	372	18.1	247	6 AAE35942	Aae35942 Lolium pe
39	343	16.7	201	3 AAG16880	Aag16880 Arabidops
40	331	16.1	190	3 AAG29671	Aag29671 Arabidops
41	321	15.6	179	3 AAG16973	Aag16973 Arabidops
42	317	15.4	216	3 AAG04466	Aag04466 Arabidops
43	314	15.3	186	3 AAG21658	Aag21658 Arabidops
44	301	14.6	179	3 AAG17410	Aag17410 Arabidops
45	292.5	14.2	226	6 AAE35947	Aae35947 Lolium pe

ALIGNMENTS

RESULT 1  
AAU12056  
ID AAU12056 standard; protein; 397 AA.  
XX  
AC AAU12056;  
XX  
DT 09-APR-2002 (first entry)  
XX  
DE Protein containing sulfotransferase encoded by Z. marina cDNA clone.  
XX  
KW Plant; transgenic; marine eelgrass; zosteric acid biosynthesis;  
KW saline-resistance; anoxia-resistance; anti-fouling genetic trait;  
KW marine vascular plant; sulphated phenolic compound; Zostera marina;  
KW sulfotransferase; S1; enzyme.  
XX  
OS Zostera marina.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 12 /label= Xaa  
FT /note= "Xaa= Stop codon, encoded by TAG"  
FT Protein 17. .347  
FT /note= "Sulfotransferase"  
FT Misc-difference 348 /label= Xaa  
FT /note= "Xaa= Stop codon, encoded by TAA"  
FT Misc-difference 354 /label= Xaa  
FT /note= "Xaa= Stop codon, encoded by TGA"  
FT Misc-difference 360 /label= Xaa  
FT /note= "Xaa= Stop codon, encoded by TGA"  
FT Misc-difference 363 /label= Xaa  
FT /note= "Xaa= Stop codon, encoded by TGA"  
FT Misc-difference 374 /label= Xaa  
FT /note= "Xaa= Stop codon, encoded by TAA"  
XX  
WO200105971-A2.  
XX  
XX 15-NOV-2001.  
XX  
XX 10-MAY-2001; 2001WO-US015412.  
XX  
XX 10-MAY-2000; 2000US-0202529P.  
XX  
XX (PHYC-) PHYCOGEN INC.  
PA

XX P1 Alberte RS, Smith RD;  
 XX DR WPI; 2002-121947/16.  
 XX DR N-PSDB; AAS20863.  
 XX PT New transgenic plants comprising a zosteric acid biosynthetic gene, a  
 XX PT saline resistance gene or a hypoxia resistance gene derived from *Zostera*  
 XX PT marina, useful for producing plants with antifouling traits.  
 XX PS Example; Fig 4; 117pp; English.  
 XX CC The present invention relates to a new transgenic plant comprising a  
 XX CC heterologous gene derived from the marine eelgrass *Zostera marina* or at  
 XX CC least one heterologous nucleotide sequence encoding a zosteric acid  
 XX CC biosynthetic function, a saline-resistance function, or a anoxia-  
 XX CC resistance function. The invention describes the method of producing a  
 XX CC transgenic plant possessing an anti-fouling genetic trait by providing a  
 XX CC cDNA population derived from a marine vascular plant, isolating from the  
 XX CC cDNA population a nucleic acid species which hybridises to a nucleic acid  
 XX CC that encodes a sulfotransferase (ST), an alcohol dehydrogenase (ADH),  
 XX CC phenylalanine ammonia lyase (PAL) or a cinnamate-4-hydroxylase (CH), and  
 XX CC transforming a target host plant with the isolated nucleic acid. The  
 XX CC plant is useful in the genetic engineering of plant species having  
 XX CC desirable genetic traits such as antifouling traits, salt and anoxia  
 XX CC resistance, and pathogen defence strategy. The expression of such  
 XX CC biosynthetic enzymes are sufficient to support the production of zosteric  
 XX CC acid and other sulphated phenolic compounds in a target plant. The  
 XX CC present sequence represents a protein containing sulfotransferase which  
 XX CC is encoded by a *Z. marina* cDNA clone  
 XX SQ Sequence 397 AA;  
 Query Match 96.9%; Score 1996; DB 5; Length 397;  
 Best Local Similarity 96.5%; Pred. No. 3.4e-166;  
 Matches 391; Conservative 0; Mismatches 0; Indels 6; Gaps 6;  
 QY 1 TRGITGIAVAC-LPLIMAGILALEKCFGSKNEQKEEDSKMYKRYREIVSSLPNDYWDG 59  
 DB 1 TRGITGIAVACXPLIMAGILALEKCFGSKNEQKEEDSKMYKRYREIVSSLPNDYWDG 60  
 QY 60 TMRLYKGFQWGVLPVGINAFENFKARETDIIITLTPKAGTTWKALTFAILLTRDVNHP 119  
 DB 61 TMRLYKGFQWGVLPVGINAFENFKARETDIIITLTPKAGTTWKALTFAILLTRDVNHP 120  
 QY 120 SSPTPLFFNPHSCVQNLLEYLYMGRENTMPDLMLNESPRLPFAGHIPYSLLPASVLKSG 179  
 DB 121 SSPTPLFFNPHSCVQNLLEYLYMGRENTMPDLMLNESPRLPFAGHIPYSLLPASVLKSG 180  
 QY 180 TKIINISRNKSTFVSFWKGNLINPDKLLDLEKSVDFISGISFCGPEWNPQAEFTNAA 239  
 DB 181 TKIINISRNKSTFVSFWKGNLINPDKLLDLEKSVDFISGISFCGPEWNPQAEFTNAA 240  
 QY 240 STNSNLLLSYEEMLEKFPVENVKKLAEFMCGCGFTDDEEKQGIIVDEIVKLCSPDNLKNQV 299  
 DB 241 STNSNLLLSYEEMLEKFPVENVKKLAEFMCGCGFTDDEEKQGIIVDEIVKLCSPDNLKNQV 300  
 QY 300 NKGSSVNSKIDNKHFRKGEVDWANYLTSEMVKLETAGINSE-KHLIS-NKULT- 356  
 DB 301 NKGSSVNSKIDNKHFRKGEVDWANYLTSEMVKLETAGINSEKXHLISXNKNUTX 360  
 QY 357 NF-NLANNYCKSN-ISLCLLSYSFVINNFILKXXXX 391  
 DB 361 NFXNLANNYCKSNISLCLLSYSFVINNFILKXXXX 397  
 RESULT 2  
 AAU12057  
 ID AAU12057 standard; protein; 331 AA.  
 XX AC AAU12057;  
 XX AC  
 DT 09-APR-2002 (first entry)

XX Zostera marina sulfotransferase protein.  
 XX DE Plant; transgenic; marine eelgrass; zosteric acid biosynthesis;  
 XX KW saline-resistance; anoxia-resistance; anti-fouling genetic trait;  
 XX KW marine vascular plant; sulphated phenolic compound; *Zostera marina*;  
 XX KW sulfotransferase; ST; enzyme.  
 XX OS *Zostera marina*.  
 XX PN WO200185971-A2.  
 XX XX 15-NOV-2001.  
 XX PD 10-MAY-2001; 2001WO-US015412.  
 XX PF 10-MAY-2000; 2000US-020529P.  
 XX PR (PHYC-) PHYCOGEN INC.  
 XX PA Alberte RS, Smith RD;  
 XX PI WPI; 2002-121947/16.  
 XX DR New transgenic plants comprising a zosteric acid biosynthetic gene, a  
 XX PT saline resistance gene or a hypoxia resistance gene derived from *Zostera*  
 XX PT marina, useful for producing plants with antifouling traits.  
 XX PS Example; Fig 5; 117pp; English.  
 XX CC The present invention relates to a new transgenic plant comprising a  
 XX CC heterologous gene derived from the marine eelgrass *Zostera marina* or at  
 XX CC least one heterologous nucleotide sequence encoding a zosteric acid  
 XX CC biosynthetic function, a saline-resistance function, or a anoxia-  
 XX CC resistance function. The invention describes the method of producing a  
 XX CC transgenic plant possessing an anti-fouling genetic trait by providing a  
 XX CC cDNA population derived from a marine vascular plant, isolating from the  
 XX CC cDNA population a nucleic acid species which hybridises to a nucleic acid  
 XX CC that encodes a sulfotransferase (ST), an alcohol dehydrogenase (ADH),  
 XX CC phenylalanine ammonia lyase (PAL) or a cinnamate-4-hydroxylase (CH), and  
 XX CC transforming a target host plant with the isolated nucleic acid. The  
 XX CC plant is useful in the genetic engineering of plant species having  
 XX CC desirable genetic traits such as antifouling traits, salt and anoxia  
 XX CC resistance, and pathogen defence strategy. The expression of such  
 XX CC biosynthetic enzymes are sufficient to support the production of zosteric  
 XX CC acid and other sulphated phenolic compounds in a target plant. The  
 XX CC present sequence represents *Z. marina* sulfotransferase protein  
 XX SQ Sequence 331 AA;  
 Query Match 85.0%; Score 1750; DB 5; Length 331;  
 Best Local Similarity 100.0%; Pred. No. 9.6e-145;  
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 16 MAGILALEKCFGSKNEQKEEDSKMYKRYREIVSSLPNDYWDGTMRLYKGFQWGVLP 75  
 DB 1 MAGILALEKCFGSKNEQKEEDSKMYKRYREIVSSLPNDYWDGTMRLYKGFQWGVLP 60  
 QY 76 GIMAFEDNFKARETDIIITLTPKAGTTWKALTFAILLTRDVNHPSSPTPLFFNPHSCV 135  
 DB 61 GIMAFEDNFKARETDIIITLTPKAGTTWKALTFAILLTRDVNHPSSPTPLFFNPHSCV 120  
 QY 136 QNLEYLYMGRENTMPDLMLNESPRLPFAGHIPYSLLPASVLKSGTKIINISRNKSTFVS 195  
 DB 121 QNLEYLYMGRENTMPDLMLNESPRLPFAGHIPYSLLPASVLKSGTKIINISRNKSTFVS 180  
 QY 196 FWKFGNLIINPDKLLDLEKSVDFISGISFCGPEWNPQAEFTNAASTNSNLLLSYEEMLE 255  
 DB 181 FWKFGNLIINPDKLLDLEKSVDFISGISFCGPEWNPQAEFTNAASTNSNLLLSYEEMLE 240  
 QY 256 KPVENVKKLAEFMCGCGFTDDEEKQGIIVDEIVKLCSPDNLKNQVNSKIDNKH 315  
 DB 241 KPVENVKKLAEFMCGCGFTDDEEKQGIIVDEIVKLCSPDNLKNQVNSKIDNKH 300

Qy 316 FRKGEVDWANYLTSEMINKKETAGKINESE 346  
Db 301 FRKGEVDWANYLTSEMINKKETAGKINESE 331

## RESULT 3

AAG21657  
ID AAG21657 standard; protein; 331 AA.

XX AC AAG21657;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 24291.

XX KW Protein identification; signal transduction pathway; metabolic pathway;  
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX KW termination sequence.

OS Arabidopsis thaliana.

XX FN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-00301439.

XX PR 25-FEB-1999; 99US-0121825P.

PR 05-MAR-1999; 99US-0123180P.

PR 09-MAR-1999; 99US-0123548P.

PR 23-MAR-1999; 99US-0125788P.

PR 25-MAR-1999; 99US-0126264P.

PR 29-MAR-1999; 99US-0126785P.

PR 01-APR-1999; 99US-0127462P.

PR 08-APR-1999; 99US-0128234P.

PR 08-APR-1999; 99US-0128714P.

PR 16-APR-1999; 99US-0129845P.

PR 19-APR-1999; 99US-0130077P.

PR 21-APR-1999; 99US-0130449P.

PR 23-APR-1999; 99US-0130510P.

PR 28-APR-1999; 99US-0130891P.

PR 30-APR-1999; 99US-0131449P.

PR 30-APR-1999; 99US-0132048P.

PR 04-MAY-1999; 99US-0132407P.

PR 05-MAY-1999; 99US-0132485P.

PR 06-MAY-1999; 99US-0132487P.

PR 07-MAY-1999; 99US-0132863P.

PR 11-MAY-1999; 99US-0134256P.

PR 14-MAY-1999; 99US-0134218P.

PR 14-MAY-1999; 99US-0134219P.

PR 14-MAY-1999; 99US-0134221P.

PR 18-MAY-1999; 99US-0134370P.

PR 19-MAY-1999; 99US-0134768P.

PR 20-MAY-1999; 99US-0134941P.

PR 21-MAY-1999; 99US-0135124P.

PR 21-MAY-1999; 99US-0135353P.

PR 24-MAY-1999; 99US-0135629P.

PR 25-MAY-1999; 99US-0136021P.

PR 27-MAY-1999; 99US-0136392P.

PR 28-MAY-1999; 99US-0136782P.

PR 01-JUN-1999; 99US-0137222P.

PR 03-JUN-1999; 99US-0137528P.

PR 04-JUN-1999; 99US-0137502P.

PR 07-JUN-1999; 99US-0137724P.

PR 08-JUN-1999; 99US-0138094P.

PR 10-JUN-1999; 99US-0138540P.

PR 10-JUN-1999; 99US-0138847P.

PR 14-JUN-1999; 99US-0139119P.

PR 16-JUN-1999; 99US-0139452P.

PR 16-JUN-1999; 99US-0139453P.

PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 23-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 02-AUG-1999; 99US-0147038P.  
PR 03-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147166P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.

PR	13-AUG-1999;	99US-0148684P.	Db	72	DIIVASFPKSGTTWKALTFALAORSKH--TSDNHPLLTNPHELVPYLELDYL--KSS	127
PR	16-AUG-1999;	99US-0149368P.	Qy	149	MPDLML-NESPRLPAGHIPYSLLPASVLKSGTKIINISNRKSTFTVSWKFGNLIINPKD	207
PR	17-AUG-1999;	99US-0149175P.	Db	128	KEDLTPLSPSPRLFTSTMSFPDALKVPLKSPCKIVYVCRNVNDVLISLWCFENSMGSEN	187
PR	20-AUG-1999;	99US-0149722P.	Qy	208	LIDLEKSVDIIFASGISFCGPFW-NQAEFTNAASTNSNLLLSYEEMLEKPVENVKLA	266
PR	20-AUG-1999;	99US-0149723P.	Db	188	NLSLEALFESLCSGYNLCPLWENVLGYWRGSLDPKHVLFRIYEELKTEPRVQIKLA	247
PR	23-AUG-1999;	99US-0149902P.	Qy	267	FMGCGFTDDEERQGIIVDEIVLKCSFDNLKNOQVKNKGSSYNKIDNKHFFRKGEVDR	326
PR	25-AUG-1999;	99US-0150566P.	Db	248	FLDCPFTKEEDSGVDKILEICSLRNLSGLEINKTG-SLSEGVSKFFRKGEVGDWKS	306
PR	26-AUG-1999;	99US-0150884P.	Qy	327	YLTSEMICKLE	337
PR	27-AUG-1999;	99US-0151065P.	Db	307	YMTPEMENKID	317
PR	27-AUG-1999;	99US-0151080P.	Qy			
PR	30-AUG-1999;	99US-0151303P.	Db			
PR	31-AUG-1999;	99US-0151438P.	Qy			
PR	01-SEP-1999;	99US-0151930P.	Db			
PR	07-SEP-1999;	99US-0152363P.	Qy			
PR	10-SEP-1999;	99US-0153070P.	Db			
PR	13-SEP-1999;	99US-0153758P.	Qy			
PR	15-SEP-1999;	99US-0154018P.	Db			
PR	16-SEP-1999;	99US-0154039P.	Qy			
PR	20-SEP-1999;	99US-0154779P.	Db			
PR	22-SEP-1999;	99US-0155133P.	Qy			
PR	23-SEP-1999;	99US-0155486P.	Db			
PR	24-SEP-1999;	99US-0155659P.	Qy			
PR	28-SEP-1999;	99US-0156458P.	Db			
PR	29-SEP-1999;	99US-0156596P.	Qy			
PR	04-OCT-1999;	99US-0157117P.	Db			
PR	05-OCT-1999;	99US-0157533P.	Qy			
PR	06-OCT-1999;	99US-0157865P.	Db			
PR	07-OCT-1999;	99US-0158029P.	Qy			
PR	08-OCT-1999;	99US-0158232P.	Db			
PR	12-OCT-1999;	99US-0158369P.	Qy			
PR	13-OCT-1999;	99US-0159293P.	Db			
PR	13-OCT-1999;	99US-0159294P.	Qy			
PR	13-OCT-1999;	99US-0159295P.	Db			
PR	14-OCT-1999;	99US-0159329P.	Qy			
PR	14-OCT-1999;	99US-0159330P.	Db			
PR	14-OCT-1999;	99US-0159331P.	Qy			
PR	14-OCT-1999;	99US-0159637P.	Db			
PR	14-OCT-1999;	99US-0159638P.	Qy			
PR	18-OCT-1999;	99US-0159584P.	Db			
PR	21-OCT-1999;	99US-0160741P.	Qy			
PR	21-OCT-1999;	99US-0160767P.	Db			
PR	21-OCT-1999;	99US-0160768P.	Qy			
PR	21-OCT-1999;	99US-0160770P.	Db			
PR	21-OCT-1999;	99US-0160814P.	Qy			
PR	21-OCT-1999;	99US-0160815P.	Db			
PR	22-OCT-1999;	99US-0160980P.	Qy			
PR	22-OCT-1999;	99US-0160981P.	Db			
PR	22-OCT-1999;	99US-0160989P.	Qy			
PR	23-OCT-1999;	99US-0161404P.	Db			
PR	23-OCT-1999;	99US-0161405P.	Qy			
PR	23-OCT-1999;	99US-0161406P.	Db			
PR	26-OCT-1999;	99US-0161359P.	Qy			
PR	26-OCT-1999;	99US-0161360P.	Db			
PR	26-OCT-1999;	99US-0161361P.	Qy			
PR	28-OCT-1999;	99US-0161920P.	Db			
PR	28-OCT-1999;	99US-0161922P.	Qy			
PR	28-OCT-1999;	99US-0161933P.	Db			
PR	29-OCT-1999;	99US-0162142P.	Qy			
Query Match 27.8%; Score 572.5; DB 3; Length 331;						
Best local similarity 39.5%; Pred. No. 1.7e-41;						
Matches 123; Conservative 61; Mismatches 118; Indels 9; Gaps 7;						
Qy	31	EOEKEESKMKYKREIVSSLPN-DYWGDTMRLYKGFQWQVIMAFEDNFKARET	89			
Db	12	EEEEEEENQSEETKSLISSLPSSIDCGTKLYKQCWDKLLQILNINFNQFQET	71			
Qy	90	DIILTLTPKAGTTWKALTFAITLRDYNHPSPTPLFFNPHSCVQNLEY-LYMGRENT	148			

RESULT 4  
AAG21656  
ID AAG21656 standard; protein; 341 AA.  
XX  
AC AAG21656;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 24290.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
FN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-00301439.  
XX  
PR 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 28-APR-1999; 99US-0130891P.  
PR 30-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.



PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 14-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140912P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 02-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142922P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145090P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145911P.  
PR 27-JUL-1999; 99US-0145912P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.

PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149923P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 28-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 28-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161320P.  
PR 28-OCT-1999; 99US-0161992P.

PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 27.8%; Score 572.5; DB 3; Length 341;  
Best Local Similarity 39.5%; Pred. No. 1.8e-41;  
Matches 123; Conservative 61; Mismatches 119; Indels 9; Gaps 7;

QY 31 POEKEEDSKMYKRYEIVSSPSN-DYWGDMRLYKGFQMGYLVPGIMAFEDNFKARET 89  
DB 22 EEEEEENQSEETKSLISSPSDIDCSGTLKYKQGCWYDKDILQALLNFKNFQPOET 81  
QY 90 DIILTLFKAGTWTKALTFAILTDVNHPSPTPLFFNPHSCVQNLVY-LYMGRENT 148  
DB 82 DIIIVAFPKSGTTLKALTFAALQSKH--TSDNHLPLTHNPHLVPYLEDLYL--KSS 137  
QY 149 MPDLMDL-NESPRLPAGHIPVSLLPASVLSKGTKIINISNRKSTFVSFWFGLNLPDK 207  
DB 138 KPDLTFLPSSSRPSTHMSFDALKVPLKESPKVIYVCRNVNDVLSLWCFEWSMSEN 197  
QY 208 LLDLEKSDVDFASGISFCQPEW-NFQAEFTNAASTNSNLLLSYEEMLEKPVENVKLA 266  
DB 198 NLSLEALFSLCSGVNLCPLWENVLYWRSLEDPKHVLFRLYBELKTEPRVQIKLAE 257  
QY 267 FMCGFTDDEEQIVDEIVKLCSDNLKNQOVNKGSSYNSKIDNKHFFRKGEVRDWA 326  
DB 258 FLDCPFTKEEDSGVDKILELCSLRLNLSGLEINKTG-SLSEGVSFKSFRRKGEVCDWKS 316  
QY 327 YLTSEMIMKLE 337  
DB 317 YMTPEMENKID 327

RESULT 5  
ABB91056  
ID ABB91056 standard; protein; 346 AA.  
AC ABB91056;  
XX ABB91056;  
DT 31-MAY-2002 (first entry)  
DE Herbicidally active polypeptide SEQ ID NO 267.  
XX Herbicidally active polypeptide; herbicide.  
XX Herbicidal; plant; agriculture; herbicide.  
XX Arabidopsis thaliana.  
XX Arabidopsis thaliana.  
XX WO200210210-A2.  
XX 07-FEB-2002.  
XX 28-AUG-2001; 2001WO-EP009892.  
XX 28-AUG-2001; 2001WO-EP009892.  
XX (FARB ) BAYER AG.  
XX Tietjen K, Weidner M;  
XX WPI; 2002-269010/31.  
XX Identifying plant target proteins for herbicidally active compounds,  
XX comprising aligning and comparing nucleic acid or amino acid sequences  
XX from plant with nucleic acid or amino acid sequences from non-plant  
XX organisms.  
XX Claim 5; SEQ ID NO 267; 261pp + Sequence Listing; English.  
XX The invention relates to identifying target proteins (ABB90790-ABB94016)  
XX for herbicidally active compounds, comprising aligning and comparing  
XX nucleic acid or amino acid sequences from plant with nucleic acid or  
XX amino acid sequences from non-plant organisms using suitable search  
XX parameters, where plant sequences having an E-value greater by a factor  
XX of 3 than the E-value of most similar non-plant sequences are selected.

CC The polypeptides or nucleic acids encoding them are useful for  
CC identifying modulators. The identified modulators are useful as  
CC herbicides  
XX  
SQ Sequence 346 AA;  
Query Match 27.5%; Score 567; DB 5; Length 346;  
Best Local Similarity 39.6%; Pred. No. 5.5e-41;  
Matches 137; Conservative 56; Mismatches 123; Indels 30; Gaps 11;

QY 23 EKCFSKNEQKEEDSKMYKRYEIVSSPSNDY--GDTWRLYKGFQMGYLVPGIMAF 80  
DB 17 ELASSPSSEFKNQ-----KHQEIATLPHKDGWRPKDPFVEYGGHWWLQPLLEGLLHA 71  
QY 81 ENFKARETDIILTLPKAGTTWTKALTFAILTR---DYNHPSSTHPLFFNPHSCVQN 137  
DB 72 QKFFKARPNDFFVCSYPKGTTLTKALTFAIANRSKFDVS-----TNPLKENPHEFPY 126  
QY 138 LEVLYMGRENTMPDLMDL-NESPRLPAGHIPVSLLPASVLSKGTKIINISNRKSTFVSF 196  
DB 127 IEIDP-----PFFPSVDVLKDEGNTLFTTHIPYDLLPESVVKSGCKIVYWRDPKOTFVSM 182  
QY 197 WKFGNLINPKD--LLDLEKSDVDFASGISFCQPEWNFQAEFTNAASTN-SNLLLSYEEM 253  
DB 183 WTFARKERSQQGFWVSIEEAFDKYCGLSAYGFLDHLVGYWKAYQANPDQILFLXYETM 242  
QY 254 LEKPVENVKLAEFMCGFTDDEEKOGIVDEIVKLCSDNLKNQOVNKGSSYNSK--I 310  
DB 243 RADPLPYKRLAEFGYGTKEEEGVNVEKVKVLCSPFETLKNLEANKGEKDRDPVY 302  
QY 311 DNKHFRKGEVRDWAANYLTSEMIMKLETAGKINESEK--HLLSNKN 354  
DB 303 ANSAYFRKGVGDWQNYLTPEMVARID--GLMBEKFKGTFGLSSKS 346

RESULT 6  
ABP72964  
ID ABP72964 standard; protein; 346 AA.  
XX ABP72964;  
XX ABP72964;  
DT 28-MAY-2003 (first entry)  
DE Amino acid sequence of desulfoglucosinolate sulfotransferase AtST5c.  
XX Plant; desulfoglucosinolate sulfotransferase; AtST5a; AtST5b; AtST5c;  
XX glucosinolate; transgenic plant.  
XX Arabidopsis thaliana.  
XX OS Arabidopsis thaliana.  
XX WO2003010318-A2.  
XX 06-FEB-2003.  
XX 24-JUL-2002; 2002WO-CA001144.  
XX 24-JUL-2001; 2001US-0307141P.  
XX (UYCO-) UNIV CONCORDIA.  
XX Varin L, Spertini D;  
XX WPI; 2003-248082/24.  
XX N-PSDB; ABZ68953.  
XX Novel isolated or purified polypeptide having biological activity of  
XX desulfoglucosinolate sulfotransferases, useful as enzymes involved in  
XX glucosinolate biosynthesis in transgenic plants.  
XX Claim 35; Page 62-63; 68pp; English.  
XX The present sequence represents a plant desulfoglucosinolate  
XX sulfotransferase, designated AtST5c. The specification also describes  
XX

CC AtST5a and AtST5b. AtST5a and AtST5b are constitutively expressed in all  
 CC parts of the plants, while AtST5c does not seem to be expressed in the  
 CC early stage of development. These enzymes are involved in glucosinolate  
 CC synthesis in plants. The desulfoglucosinolate sulfotransferase  
 CC polynucleotides are useful for modulating glucosinolate biosynthesis in  
 CC transgenic plants  
 XX Sequence 346 AA;

Query Match 27.5%; Score 567; DB 6; Length 346;  
 Best Local Similarity 39.6%; Pred. No. 5.5e-41;  
 Matches 137; Conservative 56; Mismatches 123; Indels 30; Gaps 11;  
 QY 23 EKCFGSKNEQKEEDSKMYKRYREIVSLPSNDY--GDTWRLYKGFQWQGYLVPGIMAF 80  
 Db 17 ELASSPSEFEKQ-----KHVQLIATLPHKDGWRPKDPVEYGGHWWLQPLLEGLHA 71  
 QY 81 EDNFKARTDILITLTPKAGTTWTYKALTFAILTR---DVNHPSTPLPLFFNPHSCVQN 137  
 Db 72 QKFFKARENDPFCVSYPKTGTITLTKALTFAIANRKFDS-----TNPLKRNPHFEV 126  
 QY 138 LEYLVMGRENTPDML--NESPLFAGHIPYSLLPASVLKSGTKIINISNRKSTTVSF 196  
 Db 127 IEIDF-----PFPPSDVLKDEGNTLFTSHIPYDILLPSVKSCKIYIWRDPKDTFVM 182  
 QY 197 WKFGNLIINPK--LLDLEKSYDIFASGISFCGPEWFOAEFTNAASTN--SNLLLSYEEM 253  
 Db 183 WFAHKERSQQQPVVSIIEAFKCYQGLSAYGPLYDHLVGYWKAYQANPDQILFLKYETM 242  
 QY 254 LEKPVNKKLAENFGCGFTDDEKQGIIVDEIVKLCSPDNLKQVKNKNGSSYNSK---I 310  
 Db 243 RADPLPYKRLAENFGYFTKEEBEGNVVEKVKLCSETLKNLEANKGEKDRDPVY 302  
 QY 311 DNKHFRRGEVRDWNLYLTSEMIRKLETAGKINESEK--HLLSNKN 354  
 Db 303 ANSAVFRKGVGDWQNYLTPEVARIID--GLMEKFKGTGFLSSKS 346

RESULT 7  
 ABP72962  
 ID ABP72962 standard; protein; 338 AA.  
 AC ABP72962;  
 DT 28-MAY-2003 (first entry)  
 XX  
 DE Amino acid sequence of desulfoglucosinolate sulfotransferase AtST5a.  
 XX Plant; desulfoglucosinolate sulfotransferase; AtST5a; AtST5b; AtST5c;  
 KW glucosinolate; transgenic plant.  
 XX Arabidopsis thaliana.  
 OS Arabidopsis thaliana.  
 XX WO2003010318-A2.  
 PN 06-FEB-2003.  
 XX 24-JUL-2002; 2002WO-CA0001144.  
 PF 24-JUL-2001; 2001US-0307141P.  
 PR (UYCO-) UNIV CONCORDIA.  
 XX Varin L, Spertini D;  
 XX WPI; 2003-248082/24.  
 DR N-PSDB; ABZ68951.  
 XX Novel isolated or purified polypeptide having biological activity of  
 PT desulfoglucosinolate sulfotransferases, useful as enzymes involved in  
 PT glucosinolate biosynthesis in transgenic plants.  
 XX Claim 35; Page 56-57; 68pp; English.

XX The present sequence represents a plant desulfoglucosinolate  
 CC sulfotransferase, designated AtST5a. The specification also describe  
 CC AtST5b and AtST5c. AtST5a and AtST5b are constitutively expressed in all  
 CC parts of the plants, while AtST5c does not seem to be expressed in the  
 CC early stage of development. These enzymes are involved in glucosinolate  
 CC synthesis in plants. The desulfoglucosinolate sulfotransferase  
 CC polynucleotides are useful for modulating glucosinolate biosynthesis in  
 CC transgenic plants  
 XX Sequence 338 AA;

Query Match 27.3%; Score 562.5; DB 6; Length 338;  
 Best Local Similarity 38.4%; Pred. No. 1.3e-40;  
 Matches 129; Conservative 59; Mismatches 123; Indels 25; Gaps 10;  
 QY 33 EKEEDSKMYKRYREIVSLPSNDY--GDTWRLYKGFQWQGYLVPGIMAFEDNFKARETD 90  
 Db 14 EUTEFEKTOKKYQDFIATLPHKSGWRPDEILTYQGGHWWQEBCLLEGLFHAKDHFARPTD 73  
 QY 91 IILTTLPKAGTTWTYKALTFAILTRDVPNHPSSPTPLFFNPHSCVQNLEYLYMGRNTMP 150  
 Db 74 FLVCSYPTKGTITLTKALTFAIVNR--SRVYDAAANPLKRNPHFVYVEIDFA---FYP 127  
 QY 151 DLDMLNESPR-LFAGHIPYSLLPASVLKSGTKIINISNRKSTTVSFYKFGNLIINPK-- 207  
 Db 128 TVDVLQDRKNPLFSTHINPLNGLLPDSIVNSGCKWYIWRDPKDTFISMMTF---LHKEK 184  
 QY 208 ---LLDLEKSYDIFASGISFCGPEWFOAEFTNAASTN--SNLLLSYEEMLEKPVENYK 263  
 Db 185 EGQLASLEDSDFMFCGLSVGYPLDHLVGYWKAYQENPDRLFLRYETMTMRANPLPFVKR 244  
 QY 264 LAEFMCGGTDEEKQGIIVDEIVKLCSPDNLKQVKNKNGSSYNSK---IDNKHFRKE 320  
 Db 245 LAEFMCGYGTDEEENGVAEKVKLCSETLKNLEANKGDKERDRPVPYANSAYFRK 304  
 QY 321 VRDWNLYLTSEMIRKLETAGKINESEK--HLLSNKN 354  
 Db 305 VGDWNYLTPEVARIID--GLVEEKFQDTGLLOHND 338

RESULT 8  
 AAG16972  
 ID AAG16972 standard; protein; 338 AA.  
 AC AAG16972;  
 XX 17-OCT-2000 (first entry)  
 DT Arabidopsis thaliana protein fragment SEQ ID NO: 17817.  
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 17817.  
 XX Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 XX Arabidopsis thaliana.  
 OS Arabidopsis thaliana.  
 XX EP1033405-A2.  
 XX 06-SEP-2000.  
 PD 25-FEB-2000; 2000EP-00301439.  
 PF 25-FEB-1999; 99US-0121825P.  
 PR 05-MAR-1999; 99US-0123180P.  
 PR 09-MAR-1999; 99US-0123548P.  
 PR 23-MAR-1999; 99US-0125788P.  
 PR 25-MAR-1999; 99US-0126264P.  
 PR 29-MAR-1999; 99US-0126785P.  
 PR 01-APR-1999; 99US-0127462P.  
 PR 06-APR-1999; 99US-0128234P.  
 PR 08-APR-1999; 99US-0128714P.  
 PR 16-APR-1999; 99US-0129845P.



PR	14-OCT-1999;	99US-0159330P.	XX	25-FEB-2000;	200CEP-00301439.	XX
PR	14-OCT-1999;	99US-0159331P.	PF	25-FEB-1999;	99US-0121825P.	XX
PR	14-OCT-1999;	99US-0159337P.	XX	05-MAR-1999;	99US-0123180P.	PR
PR	14-OCT-1999;	99US-0159338P.	PR	03-MAR-1999;	99US-012358P.	PR
PR	18-OCT-1999;	99US-0159384P.	PR	29-MAR-1999;	99US-0125788P.	PR
PR	21-OCT-1999;	99US-0160741P.	PR	25-MAR-1999;	99US-0126264P.	PR
PR	21-OCT-1999;	99US-0160767P.	PR	29-MAR-1999;	99US-0126785P.	PR
PR	21-OCT-1999;	99US-0160770P.	PR	01-APR-1999;	99US-0127462P.	PR
PR	21-OCT-1999;	99US-0160814P.	PR	06-APR-1999;	99US-0128234P.	PR
PR	21-OCT-1999;	99US-0160815P.	PR	08-APR-1999;	99US-0128714P.	PR
PR	21-OCT-1999;	99US-0160880P.	PR	16-APR-1999;	99US-0129845P.	PR
PR	22-OCT-1999;	99US-0160981P.	PR	19-APR-1999;	99US-0130072P.	PR
PR	22-OCT-1999;	99US-0160989P.	PR	21-APR-1999;	99US-0130449P.	PR
PR	25-OCT-1999;	99US-0161404P.	PR	23-APR-1999;	99US-0130510P.	PR
PR	25-OCT-1999;	99US-0161405P.	PR	28-APR-1999;	99US-0130891P.	PR
PR	25-OCT-1999;	99US-0161406P.	PR	28-APR-1999;	99US-0131449P.	PR
PR	26-OCT-1999;	99US-0161359P.	PR	30-APR-1999;	99US-0132048P.	PR
PR	26-OCT-1999;	99US-0161360P.	PR	30-APR-1999;	99US-0132407P.	PR
PR	26-OCT-1999;	99US-0161361P.	PR	04-MAY-1999;	99US-0132484P.	PR
PR	28-OCT-1999;	99US-0161920P.	PR	05-MAY-1999;	99US-0132485P.	PR
PR	28-OCT-1999;	99US-0161920P.	PR	06-MAY-1999;	99US-0132486P.	PR
PR	28-OCT-1999;	99US-0161933P.	PR	07-MAY-1999;	99US-0132487P.	PR
PR	29-OCT-1999;	99US-0162142P.	PR	07-MAY-1999;	99US-0132863P.	PR
Query Match 27.38; Score 561.5; DB 3; Length 338;						
Best Local Similarity 38.44; Pred. No. 1.6e-40;						
Matches 129; Conservative 59; Mismatches 123; Indels 25; Gaps 10;						
QY	33	EKEDSKMYKRYRIVSLSSNDYW--GDTWRLYKGFQWQGVLPVPGIMAFEDNFKARETD 90		11-MAY-1999;	99US-0134256P.	PR
Db	14	ELTEFEKTKYQDFIATLPKSGWRPDEILTYGGHWWQECLEGLFHAQDHFEARPTD 73		14-MAY-1999;	99US-0134218P.	PR
QY	91	IILTLTKAGTWTWKATPAILTRDVNHPSPHPLFPNPHSCVQNVLYMGRNTWP 150		14-MAY-1999;	99US-0134221P.	PR
Db	74	FLVCSYKTKGTWIKALTYAIVNR--SRYDDAANPLKRNPHFVPYVEIDFA----FYP 127		14-MAY-1999;	99US-0134370P.	PR
QY	151	DLDMLESPR-LPAGHIPYSLLPASVLKSGTKINISRNKSTFVSFKWFGNLIINPKD-- 207		18-MAY-1999;	99US-0134768P.	PR
Db	128	TVDVLQDRKPLFSTHPLNGLLPDSIVNSGCKWYIWRDPKDFISWTF---LHREKSQ 184		19-MAY-1999;	99US-0134941P.	PR
QY	208	---LLDLEKSVDIIPASIGFCGPEWFOAFTNAASTN-SNLLLSYEEMLEKPVENVKK 263		20-MAY-1999;	99US-0135124P.	PR
Db	185	EGQLASLEDSFDMPCKGLSVGYPYLDHVLGYWKAYQENPDRIILFLRYETWRANPLPFVKR 244		21-MAY-1999;	99US-0135353P.	PR
QY	264	LAEFMGCGFTDDEKQGVDEIVKLCGFDNLKNOOVNKGSSYNSK---IDNKHFFRKGE 320		24-MAY-1999;	99US-0135623P.	PR
Db	245	LAEFMGVGTDEEENGVAEKVKVLCGFETLKNLEANKGDKEREDRPAYVANSAYFRKKG 304		25-MAY-1999;	99US-0136021P.	PR
QY	321	VRDWANYLTSEMIXKLETAGKINESEKH--LLSNKN 354		27-MAY-1999;	99US-0136392P.	PR
Db	305	VGDWANYLTPEMAARID--GILEEKFKDTGLLOHDN 338		28-MAY-1999;	99US-0136782P.	PR
RESULT 9						
AAG17409						
ID	AAG17409 standard; protein; 350 AA.					
AC	AAG17409;					
XX	17-OCT-2000 (first entry)					
DT	Arabidopsis thaliana protein fragment SEQ ID NO: 18415.					
DE	Protein identification; signal transduction pathway; metabolic pathway;					
KW	hybridisation assay; genetic mapping; gene expression control; promoter;					
KW	termination sequence.					
XX	Arabidopsis thaliana.					
OS	EP1033405-A2.					
PN	06-SEP-2000.					
PD						



DT 31-MAY-2002 (first entry)  
 DE Herbicidally active polypeptide SEQ ID NO 800.  
 KW Herbicidal; plant; agriculture; herbicide.  
 XX Arabidopsis thaliana.  
 OS WO200210210-A2.  
 PN 07-FEB-2002.  
 PD 28-AUG-2001; 2001WO-EP009892.  
 PF 28-AUG-2001; 2001WO-EP009892.  
 PR (FARB) BAYER AG.  
 XX Tietjen K, Weidner M;  
 XX WPI; 2002-269010/31.  
 DR Identifying plant target proteins for herbicidally active compounds,  
 PT comprising aligning and comparing nucleic acid or amino acid sequences  
 PT from plant with nucleic acid or amino acid sequences from non-plant  
 PT organisms.  
 XX Claim 5; SEQ ID NO 800; 261pp + Sequence Listing; English.  
 PS The invention relates to identifying target proteins (ABB90790-ABB94016)  
 CC for herbicidally active compounds, comprising aligning and comparing  
 CC nucleic acid or amino acid sequences from plant with nucleic acid or  
 CC amino acid sequences from non-plant organisms using suitable search  
 CC parameters, where plant sequences having an E-value greater by a factor  
 CC of 3 than the E-value of most similar non-plant sequences are selected.  
 CC The polypeptides or nucleic acids encoding them are useful for  
 CC identifying modulators. The identified modulators are useful as  
 CC herbicides  
 XX Sequence 350 AA;  
 SQ  
 Query Match 27.1%; Score 558.5; DB 5; Length 350;  
 Best Local Similarity 39.0%; Pred. No. 3.1e-40; Indels 23; Gaps 9;  
 Matches 129; Conservative 59; Mismatches 120;  
 QY 29 KNEQKEEDSKMYKRYREIVSSLPNDYWGDTMRL--YKGFQMGYLVPGIMAFEDNPKA 86  
 DB 22 ETKTESTEPEKQKRYQDLISTFPHEKGWRPKPELIEYGGYWWLPSSLGCIHAQEFFQA 81  
 QY 87 RETDIILTLTPKAGTTWKALTFAILTRDYNHPSSTHPLLFNPHSCVQNLVLYMGRE 146  
 DB 82 RPSDFLVCSYPTGTTWLKALTFAIANRSRFDSS--NPLLRNPHFVPIEIDF--- 135  
 QY 147 NTMPDMLNESPRLFAGHIPYSLLPASVLKSGTKIINISNRKSTFVSFWKFGNINP 205  
 DB 136 PPFPEVDVLKXGNTLFTSHIPYELLPSDVVKSCKWVYIWRPKDTFISMWTF---LHK 192  
 QY 206 DK-----LLDLEKSVDFASGISFCGPEWFOAEFTNAASTN-SNLLLSYEEMLEKPE 259  
 DB 193 ERTGLGPNVSNLEESDFMFCRGLSGYGYLHILAYWKAYQENPDRILFLKYETMRADPLP 252  
 QY 260 NVKLAEFMCGFTDDEEKQIVDEIVKLCSPNLKNQOVNKGSSYNSK---IDNKHFF 316  
 DB 253 YVKSLEAFMGHGHTAEEEKGVKVVNLCSPFTLKNLEANKGEKREDRPGVYANSAYF 312  
 QY 317 RKGEVRDWANYLTSEMICKLETAGKINESEK 347  
 DB 313 RKGKVGDSNYLTPMAARID--GLMEKFK 341  
 RESULT 11  
 ABP72963  
 ID ABP72963 standard; protein; 350 AA.

XX ABP72963;  
 XX 28-MAY-2003 (first entry)  
 XX Amino acid sequence of desulfooglucosinolate sulfotransferase AtST5b.  
 DE plant; desulfooglucosinolate sulfotransferase; AtST5a; AtST5b; AtST5c;  
 XX glucosinolate; transgenic plant.  
 XX Arabidopsis thaliana.  
 OS WO2003010318-A2.  
 PN 06-FEB-2003.  
 PD 24-JUL-2002; 2002WO-CA001144.  
 PF 24-JUL-2001; 2001US-0307141P.  
 PR (UYCO-) UNIV CONCORDIA.  
 XX Varin L, Spertini D;  
 XX WPI; 2003-248082/24.  
 DR N-PSDB; ABZ68952.  
 XX Novel isolated or purified polypeptide having biological activity of  
 PT desulfooglucosinolate sulfotransferases, useful as enzymes involved in  
 PT glucosinolate biosynthesis in transgenic plants.  
 XX Claim 35; Page 59-60; 68pp; English.  
 CC The present sequence represents a plant desulfooglucosinolate  
 CC sulfotransferase, designated AtST5b. The specification also describes  
 CC AtST5a and AtST5c. AtST5a and AtST5b are constitutively expressed in all  
 CC parts of the plants, while AtST5c does not seem to be expressed in the  
 CC early stage of development. These enzymes are involved in glucosinolate  
 CC synthesis in plants. The desulfooglucosinolate sulfotransferase  
 CC polynucleotides are useful for modulating glucosinolate biosynthesis in  
 CC transgenic plants  
 XX Sequence 350 AA;  
 SQ  
 Query Match 27.1%; Score 558.5; DB 6; Length 350;  
 Best Local Similarity 39.0%; Pred. No. 3.1e-40; Indels 23; Gaps 9;  
 Matches 129; Conservative 59; Mismatches 120;  
 QY 29 KNEQKEEDSKMYKRYREIVSSLPNDYWGDTMRL--YKGFQMGYLVPGIMAFEDNPKA 86  
 DB 22 ETKTESTEPEKQKRYQDLISTFPHEKGWRPKPELIEYGGYWWLPSSLGCIHAQEFFQA 81  
 QY 87 RETDIILTLTPKAGTTWKALTFAILTRDYNHPSSTHPLLFNPHSCVQNLVLYMGRE 146  
 DB 82 RPSDFLVCSYPTGTTWLKALTFAIANRSRFDSS--NPLLRNPHFVPIEIDF--- 135  
 QY 147 NTMPDMLNESPRLFAGHIPYSLLPASVLKSGTKIINISNRKSTFVSFWKFGNINP 205  
 DB 136 PPFPEVDVLKXGNTLFTSHIPYELLPSDVVKSCKWVYIWRPKDTFISMWTF---LHK 192  
 QY 206 DK-----LLDLEKSVDFASGISFCGPEWFOAEFTNAASTN-SNLLLSYEEMLEKPE 259  
 DB 193 ERTGLGPNVSNLEESDFMFCRGLSGYGYLHILAYWKAYQENPDRILFLKYETMRADPLP 252  
 QY 260 NVKLAEFMCGFTDDEEKQIVDEIVKLCSPNLKNQOVNKGSSYNSK---IDNKHFF 316  
 DB 253 YVKSLEAFMGHGHTAEEEKGVKVVNLCSPFTLKNLEANKGEKREDRPGVYANSAYF 312  
 QY 317 RKGEVRDWANYLTSEMICKLETAGKINESEK 347  
 DB 313 RKGKVGDSNYLTPMAARID--GLMEKFK 341

RESULT 12  
AAG46836  
ID AAG46836 standard; protein; 359 AA.  
XX AC AAG46836;  
XX DT 18-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 58965.  
XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; Genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX OS Arabidopsis thaliana.  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-00301439.  
XX PR 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130443P.  
PR 23-APR-1999; 99US-0130510P.  
PR 28-APR-1999; 99US-0130891P.  
PR 30-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132040P.  
PR 04-MAY-1999; 99US-0132407P.  
PR 05-MAY-1999; 99US-0132484P.  
PR 06-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 07-MAY-1999; 99US-0132487P.  
PR 11-MAY-1999; 99US-0132863P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 18-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134944P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 17-JUN-1999; 99US-0139453P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 22-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145090P.  
PR 22-JUL-1999; 99US-0145132P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147933P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 11-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.



PR	20-AUG-1999;	99US-0149929P.	
PR	23-AUG-1999;	99US-0149902P.	
PR	23-AUG-1999;	99US-0149930P.	
PR	25-AUG-1999;	99US-0150566P.	
PR	26-AUG-1999;	99US-0150884P.	
PR	27-AUG-1999;	99US-0151065P.	
PR	27-AUG-1999;	99US-0151066P.	
PR	30-AUG-1999;	99US-0151080P.	
PR	30-AUG-1999;	99US-0151303P.	
PR	31-AUG-1999;	99US-0151438P.	
PR	01-SEP-1999;	99US-0151930P.	
PR	07-SEP-1999;	99US-0152363P.	
PR	10-SEP-1999;	99US-0153070P.	
PR	13-SEP-1999;	99US-0153758P.	
PR	15-SEP-1999;	99US-0154018P.	
PR	16-SEP-1999;	99US-0154039P.	
PR	20-SEP-1999;	99US-0154779P.	
PR	22-SEP-1999;	99US-0155139P.	
PR	23-SEP-1999;	99US-0155486P.	
PR	24-SEP-1999;	99US-0155659P.	
PR	28-SEP-1999;	99US-0156458P.	
PR	29-SEP-1999;	99US-0156596P.	
PR	04-OCT-1999;	99US-0157117P.	
PR	05-OCT-1999;	99US-0157753P.	
PR	06-OCT-1999;	99US-0157865P.	
PR	07-OCT-1999;	99US-0158029P.	
PR	08-OCT-1999;	99US-0158232P.	
PR	12-OCT-1999;	99US-0158369P.	
PR	13-OCT-1999;	99US-0159293P.	
PR	13-OCT-1999;	99US-0159294P.	
PR	13-OCT-1999;	99US-0159295P.	
PR	14-OCT-1999;	99US-0159329P.	
PR	14-OCT-1999;	99US-0159330P.	
PR	14-OCT-1999;	99US-0159331P.	
PR	14-OCT-1999;	99US-0159637P.	
PR	18-OCT-1999;	99US-0159638P.	
PR	21-OCT-1999;	99US-0159584P.	
PR	21-OCT-1999;	99US-0160741P.	
PR	21-OCT-1999;	99US-0160767P.	
PR	21-OCT-1999;	99US-0160768P.	
PR	21-OCT-1999;	99US-0160770P.	
PR	21-OCT-1999;	99US-0160814P.	
PR	21-OCT-1999;	99US-0160815P.	
PR	22-OCT-1999;	99US-0160980P.	
PR	22-OCT-1999;	99US-0160981P.	
PR	22-OCT-1999;	99US-0160989P.	
PR	25-OCT-1999;	99US-0161404P.	
PR	25-OCT-1999;	99US-0161405P.	
PR	25-OCT-1999;	99US-0161406P.	
PR	26-OCT-1999;	99US-0161359P.	
PR	26-OCT-1999;	99US-0161360P.	
PR	28-OCT-1999;	99US-0161361P.	
PR	28-OCT-1999;	99US-0161920P.	
PR	28-OCT-1999;	99US-0161992P.	
PR	28-OCT-1999;	99US-0161993P.	
PR	29-OCT-1999;	99US-0162142P.	

Query Match

Best Local Similarity 27.0%; Score 555.5; DB 3; Length 359;

Matches 123; Conservative 69; Mismatches 140; Indels 11; Gaps 8;

QY	1	TRGITGIVACUPLINAGLILEKFCGSKNEQKEEDSKMYKRYREIVSSLPNSDYGDT	60
DB	3	TSMSKSIPIAIFSFMSCHLELLKEGKTRDVPKABEDELGLSCFQEMLDLSLPKRGWTR	62
QY	61	-KRLYKGFQWQVLPVGINAFEDNFKARETDIITLPKAGTTWTWKALTFAITLRDVNHP	119
DB	63	YVLFQGFQWQAKETQAIMSKFKQSLQSLNDVVLATIPKSGTTWKALTFTILNHRDTP	122
QY	120	--SSPTHTLLFFNPHSCVQNLLEY-LYMGRENTWPDMLNESPRLFAGHIPVSLLPASVL	176
DB	123	VASSTNHPFTSNPHDLVPFFEYKLYANGD--VPDLSGL-ASPRTFATHLPFGSLKKTIE	179

QY	177	KSGTKIINISNRKSTFVSFWKFGNLINPKL--LDLEKSVDFIFASGISFCQPEWFOAE	234
DB	180	KPGVKVYLCRNPFDTFISSWHYTNNIKSESVPVLLDQAFDLYCRGVIGFPGFWEHMLG	239
QY	235	P-TNAASTNSNLLLSYEMLEKPEVNVKKLAEFWGCCTDDEEKOGIVDEIVKLCSPDN	293
DB	240	YWRESLKRPEKVFYRYEDLKDDITNLRKLTATFLELFTFEERKGVVKAIAELCSPEN	299
QY	294	LKNQOVNKGSSSYNSKIDNKHFFRKGEVRDWNANYLTSEMIKKL	336
DB	300	LKKLEVNKSNKSIKN-FENRFLFRKGEVSDWVNYLSPSQVERL	341

RESULT 13

AAG44205

ID AAG44205 standard; protein; 359 AA.

XX AAG44205;

AC AAG44205;

XX 18-OCT-2000 (first entry)

DT Arabidopsis thaliana protein fragment SEQ ID NO: 55343.

DE Arabidopsis thaliana protein fragment SEQ ID NO: 55343.

XX Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-00301439.

XX 25-FEB-1999; 99US-0121825P.

PR 05-MAR-1999; 99US-0123180P.

PR 09-MAR-1999; 99US-0123548P.

PR 23-MAR-1999; 99US-0125788P.

PR 25-MAR-1999; 99US-0126264P.

PR 23-MAR-1999; 99US-0126785P.

PR 01-APR-1999; 99US-0127462P.

PR 06-APR-1999; 99US-0128234P.

PR 08-APR-1999; 99US-0128714P.

PR 16-APR-1999; 99US-0129845P.

PR 19-APR-1999; 99US-0130077P.

PR 21-APR-1999; 99US-0130449P.

PR 23-APR-1999; 99US-0130510P.

PR 23-APR-1999; 99US-0130891P.

PR 28-APR-1999; 99US-0131449P.

PR 30-APR-1999; 99US-0132048P.

PR 04-MAY-1999; 99US-0132407P.

PR 05-MAY-1999; 99US-0132484P.

PR 06-MAY-1999; 99US-0132485P.

PR 06-MAY-1999; 99US-0132486P.

PR 06-MAY-1999; 99US-0132487P.

PR 07-MAY-1999; 99US-0132863P.

PR 11-MAY-1999; 99US-0134256P.

PR 14-MAY-1999; 99US-0134218P.

PR 14-MAY-1999; 99US-0134219P.

PR 14-MAY-1999; 99US-0134221P.

PR 14-MAY-1999; 99US-0134370P.

PR 18-MAY-1999; 99US-0134768P.

PR 19-MAY-1999; 99US-0134941P.

PR 20-MAY-1999; 99US-0135124P.

PR 21-MAY-1999; 99US-0135353P.

PR 24-MAY-1999; 99US-0135629P.

PR 25-MAY-1999; 99US-0136021P.

PR 27-MAY-1999; 99US-0136392P.

PR 28-MAY-1999; 99US-0136782P.

PR 01-JUN-1999; 99US-0137222P.

PR 03-JUN-1999; 99US-0137528P.

PR 04-JUN-1999; 99US-0137502P.





PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 06-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 24-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140931P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142300P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0130891P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145313P.  
PR 27-JUL-1999; 99US-0145318P.  
PR 27-JUL-1999; 99US-0145319P.  
PR 28-JUL-1999; 99US-0145351P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 05-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 06-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 09-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 23-AUG-1999; 99US-0149923P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 25-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155639P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157533P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158023P.  
PR 08-OCT-1999; 99US-0158234P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 14-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.

```
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 25-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161408P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match      26.8%; Score 551.5; DB 3; Length 354;
Best Local Similarity 36.2%; Pred. No. 1.3e-39;
Matches 122; Conservative 68; Mismatches 136; Indels 11; Gaps 8;

QY 7 IAVACLPITMAGILALEKCFGSKNEQKEEDSMYKRYREIVSSLPSPNDYMGDT-MRLYK 65
DB 4 IPWAI PPSFNCHKLELLKEGKTRDVPKAEDEGLSCFEQEMLDSPKRGWRTYLYLFQ 63

QY 66 GFWMQGYLVFGIMAFEDNFKARETDIITLTPKAGTTWKALTFAILTRDVNHP--SSPT 123
DB 64 GFWCQAKEIQAIMSFQKHFSLENDVLATIPKSGTTWLKALTFTILNRHRFPVASSTN 123

QY 124 HPLLFNPHSCVQNLEY-LYMGRENTMPDLDMLNESPRLFAGHIPYSLLPASVLKSGTKI 182
DB 124 HELFTSNPHDLVPFFFEYKLYANGD--VPDLSGL-ASPRTFATHLPFGSLKETIERPGVKV 180

QY 183 INISNRKSTFVSFMKFGNLINPKL--LDLEKSVDDIFASGISFCGPENWFOAEF-TNAA 239
DB 181 VVLCRNPFDTFSSWHYTNNIKSESVPVLLDQAFDLYCRGVIGFGPFWEHMLGYWRESL 240

QY 240 STNSNLLLSYEEMLEKPVENVKLAEFNGCGFTDDEEKQGIIVDEIVKLCSPDNLKNOOV 299
DB 241 KRPEKVFFLYREDLKDDDIETNLKRLATFLELPFTBEEERKGVVKAIBLCSPENLKULEV 300

QY 300 NKNGSSYNGKIDNKHFFRKGEYRDWANYLTSEMIKKL 336
DB 301 NKSNSIKN-FENRFLFRKGEVSDWNYLSPQOVERL 336
```

Search completed: March 20, 2004, 01:37:46  
Job time : 49 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2004, 01:36:08 ; Search time 17 Seconds  
(without alignments)  
1187.397 Million cell updates/sec

Title: US-09-854-122-16

Perfect score: 2059

Sequence: 1 TRGITGIAVACLPLIMAGIL.....CLLSYFVINNFILLKKKK 391

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCITUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	268	13.0	304	4	US-09-609-816-6
2	264	12.8	304	4	US-09-609-816-4
3	260	12.6	304	4	US-09-609-816-5
4	255.5	12.4	283	4	US-09-609-816-8
5	255.5	12.4	283	4	US-09-609-816-9
6	255	12.4	294	1	US-08-325-562-2
7	255	12.4	294	1	US-08-437-795-2
8	246.5	12.0	283	4	US-09-609-816-10
9	245.5	11.9	295	3	US-09-150-133-13
10	245.5	11.9	295	3	US-09-150-141-13
11	245.5	11.9	295	3	US-09-374-493-13
12	245.5	11.9	295	3	US-09-374-824-13
13	245.5	11.9	295	3	US-09-374-492-13
14	245.5	11.9	295	4	US-09-785-343-13
15	239	11.6	265	4	US-09-795-926-6
16	239	11.6	303	4	US-09-795-926-2
17	226.5	11.0	295	4	US-09-328-174A-3
18	199.5	9.7	350	2	US-08-852-481-2
19	187	9.1	284	4	US-09-786-240-7
20	184	8.9	214	4	US-09-795-926-16
21	155.5	7.6	148	4	US-09-795-926-8
22	155.5	7.6	186	4	US-09-795-926-12
23	111	5.4	110	4	US-09-795-926-4
24	111	5.4	148	4	US-09-795-926-10
25	111	5.4	1147	3	US-08-470-260-5
26	111	5.4	1147	3	US-08-471-491-5
27	111	5.4	1147	3	US-08-466-662-5

28 111 5.4 3289 2 US-08-477-451-2 Sequence 2, Appli  
29 103 5.0 859 1 US-08-053-614-2 Sequence 2, Appli  
30 103 5.0 859 1 US-08-316-397B-2 Sequence 2, Appli  
31 103 5.0 859 2 US-09-034-306-2 Sequence 2, Appli  
32 103 5.0 859 3 US-09-259-437-2 Sequence 2, Appli  
33 103 5.0 859 5 PCT-US93-09782-2 Sequence 2, Appli  
34 103 5.0 1181 1 US-08-053-614-4 Sequence 4, Appli  
35 103 5.0 1181 1 US-08-316-397B-4 Sequence 4, Appli  
36 103 5.0 1181 2 US-09-034-306-4 Sequence 4, Appli  
37 103 5.0 1181 3 US-09-259-437-4 Sequence 4, Appli  
38 103 5.0 1181 5 PCT-US93-09782-4 Sequence 4, Appli  
39 103 5.0 1467 4 US-09-134-000C-6740 Sequence 6740, Ap  
40 100.5 4.9 97 4 US-09-795-926-18 Sequence 18, Appli  
41 99.5 4.8 717 4 US-09-509-800-2 Sequence 2, Appli  
42 99 4.8 389 4 US-09-543-681A-7496 Sequence 7496, Ap  
43 98.5 4.8 1093 3 US-09-315-793-52 Sequence 52, Appli  
44 98 4.8 369 4 US-09-107-532A-5754 Sequence 5754, Ap  
45 97.5 4.7 840 3 US-08-974-549A-190 Sequence 190, App

#### ALIGNMENTS

##### RESULT 1

US-09-609-816-6  
; Sequence 6, Application US/09609816  
; Patent No. 6436684  
; GENERAL INFORMATION:  
; APPLICANT: Woodage, Trevor  
; APPLICANT: Wei, Minh Hui  
; APPLICANT: Kodira, Chinappa  
; APPLICANT: Beasley, Ellen  
; APPLICANT: DiFrancesco, Valentina  
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING  
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN  
; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF  
; FILE REFERENCE: CL000669PCT  
; CURRENT APPLICATION NUMBER: US/09/609,816  
; CURRENT FILING DATE: 2000-07-03  
; PRIOR APPLICATION NUMBER: 60/192,408  
; PRIOR FILING DATE: 2000-03-27  
; PRIOR APPLICATION NUMBER: 60/212,725  
; PRIOR FILING DATE: 2000-06-20  
; PRIOR APPLICATION NUMBER: 09/609,816  
; PRIOR FILING DATE: 2000-07-03  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 304  
; TYPE: PRT  
; ORGANISM: HUMAN  
US-09-609-816-6

Query Match 13.0%; Score 268; DB 4; Length 304;  
Best Local Similarity 27.9%; Pred. No. 3.5e-17;  
Matches 97; Conservative 56; Mismatches 131; Indels 64; Gaps 18;  
QY 29 KNEQKEESKMKYREIVTSLSPLNDYWDTRLYKGFQMGYLVPGIMAFEDNFKARE 88  
Db 6 KNAPTWEKPELFINME--VDGVP-----TLLSKWEWK-----VCNFOAP 46  
QY 89 TDILTTLPKAGTTWKALTFAILT-RDYNH----PSSETHPLLPFN-PHSCVQNLLEYL 142  
Db 47 DDILATYKSGTWTWHEILMDLNDGDYCKCKRAQTLORHAFLEKLPFKKPKDLEFV- 105  
QY 143 MGRNTWPDMLDNESPLFAGHIPVSLPASVLKSGTKIINISNRKSTFFVFWKFGNL 202  
Db 106 -----LEM--SSPOLKTHPLPSHLIPPIKENCNKIVYVARNPKDCLVSYHFHFM 154  
QY 203 IN--PKLLDLEKSVDFISGIFSCPEWNFQAEFTNAASTNSNLLLSYEENLEKPVEN 260  
Db 155 ASFWPDP-QNLEEFYKFMSC-KVCGSGFDHVKGWAAKDRHILLYFYEDIKNPKHE 212

QY 261 VKKLAFFMCGGTTDEEKGQIVDEIVKLCSPDNLNKQVNGKNGSSYNSKIDN---KHFFR 317  
DB 213 IHKVLFEFLKUSGD-----VINKLVHHTSFDMKDNPM-AHNTAVPAHFNHSISKEMR 266  
QY 318 KGEVRDWANYLTSEMIMKLETAGKINES-EKHLNKNLNTFNLNYYC 364  
DB 267 KGMFGDKNHTVAM-----NENPDKHY--EKKMAGSTL-NFC 301

RESULT 2  
US-09-609-816-4  
; Sequence 4, Application US/09609816  
; Patent No. 6436684  
; GENERAL INFORMATION:  
; APPLICANT: Woodage, Trevor  
; APPLICANT: Wei, Minh Hui  
; APPLICANT: Kodira, Chinappa  
; APPLICANT: Beasley, Ellen  
; APPLICANT: DiFrancesco, Valentina  
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING  
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN  
; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF  
; FILE REFERENCE: CL000669PCT  
; CURRENT APPLICATION NUMBER: US/09/609,816  
; CURRENT FILING DATE: 2000-07-03  
; PRIOR APPLICATION NUMBER: 60/192,408  
; PRIOR FILING DATE: 2000-03-27  
; PRIOR APPLICATION NUMBER: 60/212,725  
; PRIOR FILING DATE: 2000-06-20  
; PRIOR APPLICATION NUMBER: 09/609,816  
; PRIOR FILING DATE: 2000-07-03  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 304  
; TYPE: PRT  
; ORGANISM: HUMAN  
US-09-609-816-4

Query Match 12.8%; Score 264; DB 4; Length 304;  
Best Local Similarity 27.6%; Pred. No. 8.3e-17;  
Matches 96; Conservative 58; Mismatches 130; Indels 64; Gaps 18;

QY 29 KNEQKEEDSKMYKRYREIVSSLPNSDNDYGDWTMLRYKGFQWQGYLVFGIMAFEDNFKARE 88  
DB 6 KNAFTMEKKPELFNIME--VDGVP-----TLILSKEWWEK-----VCNFOAKP 46  
QY 89 TDIILTLPKAGTTWTKALTFAILT-RDVNH-----PSSPTHPLIFFN-PHSCVQNLEYLY 142  
DB 47 DDLILATVPKSGTTWMEHILDMILNDGVEKCKRAQTLDRHAFLEKFPKKEPDLEFV- 105  
QY 143 MGRNTMPDLDMLNESPRLPAGHIPYSLLPASVLKSGTKIINISRNKRKSTFVSFWKFGNL 202  
DB 106 -----LEM--SSQLIKTHLPShLIPPSIWKENCKIIVYVARNPKDCLVSYHFFRM 154  
QY 203 IN--PDKLLDLEKSDVIFASGISFCGPEWNPQAEFTNAASTNSNLLLSYEEMLEKPVEN 260  
DB 155 ASFPMDP-QNLEEFYEFKFMG-KVVGGSWFDHVKGWAAKDMHRILYLFYEDIKKDPKRE 212  
QY 261 VKKLAFFMCGGTTDEEKGQIVDEIVKLCSPDNLNKQVNGKNGSSYNSKIDN---KHFFR 317  
DB 213 IHKVLFEFLKUSGD-----VINKLVHHTSFDMKDNPM-AHNTAVPAHFNHSISKEMR 266  
QY 318 KGEVRDWANYLTSEMIMKLETAGKINES-EKHLNKNLNTFNLNYYC 364  
DB 267 KGMFGDKNHTVAM-----NENPDKHY--EKKMAGSTL-NFC 301

RESULT 3  
US-09-609-816-5  
; Sequence 5, Application US/09609816  
; Patent No. 6436684  
; GENERAL INFORMATION:  
; APPLICANT: Woodage, Trevor  
; APPLICANT: Wei, Minh Hui  
; APPLICANT: Kodira, Chinappa  
; APPLICANT: Beasley, Ellen  
; APPLICANT: DiFrancesco, Valentina  
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING  
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN  
; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF  
; FILE REFERENCE: CL000669PCT  
; CURRENT APPLICATION NUMBER: US/09/609,816  
; CURRENT FILING DATE: 2000-07-03  
; PRIOR APPLICATION NUMBER: 60/192,408  
; PRIOR FILING DATE: 2000-03-27  
; PRIOR APPLICATION NUMBER: 60/212,725  
; PRIOR FILING DATE: 2000-06-20  
; PRIOR APPLICATION NUMBER: 09/609,816  
; PRIOR FILING DATE: 2000-07-03  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 304  
; TYPE: PRT  
; ORGANISM: HUMAN  
US-09-609-816-5

Query Match 12.8%; Score 264; DB 4; Length 304;  
Best Local Similarity 27.6%; Pred. No. 8.3e-17;  
Matches 96; Conservative 58; Mismatches 130; Indels 64; Gaps 18;

QY 29 KNEQKEEDSKMYKRYREIVSSLPNSDNDYGDWTMLRYKGFQWQGYLVFGIMAFEDNFKARE 88  
DB 6 KNAFTMEKKPELFNIME--VDGVP-----TLILSKEWWEK-----VCNFOAKP 46  
QY 89 TDIILTLPKAGTTWTKALTFAILT-RDVNH-----PSSPTHPLIFFN-PHSCVQNLEYLY 142  
DB 47 DDLILATVPKSGTTWMEHILDMILNDGVEKCKRAQTLDRHAFLEKFPKKEPDLEFV- 105  
QY 143 MGRNTMPDLDMLNESPRLPAGHIPYSLLPASVLKSGTKIINISRNKRKSTFVSFWKFGNL 202  
DB 106 -----LEM--SSQLIKTHLPShLIPPSIWKENCKIIVYVARNPKDCLVSYHFFRM 154  
QY 203 IN--PDKLLDLEKSDVIFASGISFCGPEWNPQAEFTNAASTNSNLLLSYEEMLEKPVEN 260  
DB 155 ASFPMDP-QNLEEFYEFKFMG-KVVGGSWFDHVKGWAAKDMHRILYLFYEDIKKDPKRE 212  
QY 261 VKKLAFFMCGGTTDEEKGQIVDEIVKLCSPDNLNKQVNGKNGSSYNSKIDN---KHFFR 317  
DB 213 IHKVLFEFLKUSGD-----VINKLVHHTSFDMKDNPM-AHNTAVPAHFNHSISKEMR 266  
QY 318 KGEVRDWANYLTSEMIMKLETAGKINES-EKHLNKNLNTFNLNYYC 364  
DB 267 KGMFGDKNHTVAM-----NENPDKHY--EKKMAGSTL-NFC 301

QY 261 VKKLAFFMCGGTTDEEKGQIVDEIVKLCSPDNLNKQVNGKNGSSYNSKIDN---KHFFR 317  
DB 213 IHKVLFEFLKUSGD-----VINKLVHHTSFDMKDNPM-AHNTAVPAHFNHSISKEMR 266  
QY 318 KGEVRDWANYLTSEMIMKLETAGKINES-EKHLNKNLNTFNLNYYC 364  
DB 267 KGMFGDKNHTVAM-----NENPDKHY--EKKMAGSTL-NFC 301

RESULT 4  
US-09-609-816-8  
; Sequence 8, Application US/09609816  
; Patent No. 6436684  
; GENERAL INFORMATION:  
; APPLICANT: Woodage, Trevor  
; APPLICANT: Wei, Minh Hui  
; APPLICANT: Kodira, Chinappa  
; APPLICANT: Beasley, Ellen  
; APPLICANT: DiFrancesco, Valentina  
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING  
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN  
; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF  
; FILE REFERENCE: CL000669PCT  
; CURRENT APPLICATION NUMBER: US/09/609,816  
; CURRENT FILING DATE: 2000-07-03  
; PRIOR APPLICATION NUMBER: 60/192,408  
; PRIOR FILING DATE: 2000-03-27  
; PRIOR APPLICATION NUMBER: 60/212,725  
; PRIOR FILING DATE: 2000-06-20  
; PRIOR APPLICATION NUMBER: 09/609,816  
; PRIOR FILING DATE: 2000-07-03  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 304  
; TYPE: PRT  
; ORGANISM: HUMAN  
US-09-609-816-8

Query Match 12.6%; Score 260; DB 4; Length 304;  
Best Local Similarity 27.7%; Pred. No. 2e-16;  
Matches 86; Conservative 54; Mismatches 123; Indels 48; Gaps 13;

QY 29 KNEQKEEDSKMYKRYREIVSSLPNSDNDYGDWTMLRYKGFQWQGYLVFGIMAFEDNFKARE 88  
DB 6 KNAFTMEKKPELFNIME--VDGVP-----TLILSKEWWEK-----VCNFOAKP 46  
QY 89 TDIILTLPKAGTTWTKALTFAILT-RDVNH-----PSSPTHPLIFFN-PHSCVQNLEYLY 142  
DB 47 DDLILATVPKSGTTWMEHILDMILNDGVEKCKRAQTLDRHAFLEKFPKKEPDLEFV- 105  
QY 143 MGRNTMPDLDMLNESPRLPAGHIPYSLLPASVLKSGTKIINISRNKRKSTFVSFWKFGNL 202  
DB 106 -----LEM--SSQLIKTHLPShLIPPSIWKENCKIIVYVARNPKDCLVSYHFFRM 154  
QY 203 IN--PDKLLDLEKSDVIFASGISFCGPEWNPQAEFTNAASTNSNLLLSYEEMLEKPVEN 260  
DB 155 ASFPMDP-QNLEEFYEFKFMG-KVVGGSWFDHVKGWAAKDMHRILYLFYEDIKKDPKRE 212  
QY 261 VKKLAFFMCGGTTDEEKGQIVDEIVKLCSPDNLNKQVNGKNGSSYNSKIDN---KHFFR 318  
DB 213 IHKVLFEFLKUSGD-----VINKLVHHTSFDMKDNPM-AHNTAVPAHFNHSISKEMR 267  
QY 319 KGEVRDWANYLT 329  
DB 268 GMPGDWKNYFT 278

RESULT 4  
US-09-609-816-8  
; Sequence 8, Application US/09609816  
; Patent No. 6436684  
; GENERAL INFORMATION:  
; APPLICANT: Woodage, Trevor  
; APPLICANT: Wei, Minh Hui  
; APPLICANT: Kodira, Chinappa  
; APPLICANT: Beasley, Ellen  
; APPLICANT: DiFrancesco, Valentina  
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING  
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN  
; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF  
; FILE REFERENCE: CL000669PCT  
; CURRENT APPLICATION NUMBER: US/09/609,816  
; CURRENT FILING DATE: 2000-07-03  
; PRIOR APPLICATION NUMBER: 60/192,408  
; PRIOR FILING DATE: 2000-03-27  
; PRIOR APPLICATION NUMBER: 60/212,725  
; PRIOR FILING DATE: 2000-06-20  
; PRIOR APPLICATION NUMBER: 09/609,816  
; PRIOR FILING DATE: 2000-07-03  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 304  
; TYPE: PRT  
; ORGANISM: HUMAN  
US-09-609-816-8

```
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/609,816
; PRIOR FILING DATE: 2000-07-03
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 283
; TYPE: PRT
; ORGANISM: HUMAN
US-09-609-816-8

Query Match      12.4%; Score 255.5; DB 4; Length 283;
Best Local Similarity 26.0%; Pred. No. 4.8e-16;
Matches 76; Conservative 56; Mismatches 109; Indels 51; Gaps 10;

QY 83 NFKARETDIIILTPKAGTTWTAKLTFAI-----LTRDVNHPSPSPHLLFFNPHSCVQN 137
DB 20 NFAQKPDLLIATYAKAGTTWTQEI VDMIQNDGVQKQCRANTYDRHPFI-----69

QY 138 LEYLVMGRENTPD-----LDMLE--SPRLFAGHIPYSLLPASVLKSGTKIINSRNK 190
DB 70 -----EWTLPSPNSGLDLANKMPSPTLTKLTPVHMLPSPFWKENSIIYVARNAK 121

QY 191 STFVSFWKGNLNPDKL-----DLKSYDIFASGISFCGPEWNFQAEFTNAASTNSLL 246
DB 122 DCLVSYIYFSRM---NKMPLDPOTLGEYIEQKAGKVLWG-SWYDHVKGWWDVKQHRIL 177

QY 247 LLSYEEMLEKPVENVKLAEFMCGFTDDEEKQGIYDEIVKLCSPDNLKQVKNKNGSSY 306
DB 178 YLFYEDMKEDPKREIKKIAKFL-----EKDISSEVLNKIIYHTSFDVMKENPMANYTTL 232

QY 307 NSKIDN--KHFFKGEVRDWNVLTSEMIMKLETAGKINSEKHLSSKNLT 356
DB 233 SSIMDHSISGPFMRKMGPDWKNYFT-----VAQSEDFDEYRRKMGAGSNIT 278

RESULT 5
US-09-609-816-9
; Sequence 9, Application US/09609816
; Patent No. 643684
; GENERAL INFORMATION:
; APPLICANT: Woodage, Trevor
; APPLICANT: Wei, Minh Hui
; APPLICANT: Kodira, Chinappa
; APPLICANT: Beasley, Ellen
; APPLICANT: DiFrancesco, Valentina
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; FILE REFERENCE: CL000669PCT
; CURRENT APPLICATION NUMBER: US/09/609,816
; CURRENT FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/192,408
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: 60/212,725
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/609,816
; PRIOR FILING DATE: 2000-07-03
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 283
; TYPE: PRT
; ORGANISM: HUMAN
US-09-609-816-9

Query Match      12.4%; Score 255.5; DB 4; Length 283;
Best Local Similarity 26.0%; Pred. No. 4.8e-16;
Matches 76; Conservative 56; Mismatches 109; Indels 51; Gaps 10;

QY 83 NFKARETDIIILTPKAGTTWTAKLTFAI-----LTRDVNHPSPSPHLLFFNPHSCVQN 137
DB 20 NFAQKPDLLIATYAKAGTTWTQEI VDMIQNDGVQKQCRANTYDRHPFI-----69
```

```
QY 138 LEYLVMGRENTPD-----LDMLE--SPRLFAGHIPYSLLPASVLKSGTKIINSRNK 190
DB 70 -----EWTLPSPNSGLDLANKMPSPTLTKLTPVHMLPSPFWKENSIIYVARNAK 121

QY 191 STFVSFWKGNLNPDKL-----DLKSYDIFASGISFCGPEWNFQAEFTNAASTNSLL 246
DB 122 DCLVSYIYFSRM---NKMPLDPOTLGEYIEQKAGKVLWG-SWYDHVKGWWDVKQHRIL 177

QY 247 LLSYEEMLEKPVENVKLAEFMCGFTDDEEKQGIYDEIVKLCSPDNLKQVKNKNGSSY 306
DB 178 YLFYEDMKEDPKREIKKIAKFL-----EKDISSEVLNKIIYHTSFDVMKENPMANYTTL 232

QY 307 NSKIDN--KHFFKGEVRDWNVLTSEMIMKLETAGKINSEKHLSSKNLT 356
DB 233 SSIMDHSISGPFMRKMGPDWKNYFT-----VAQSEDFDEYRRKMGAGSNIT 278

RESULT 6
US-08-325-562-2
; Sequence 2, Application US/08325562
; Patent No. 5714594
; GENERAL INFORMATION:
; APPLICANT: Weinshilboum, Richard M.
; APPLICANT: Aksoy, Ibrahim A.
; APPLICANT: Wood, Thomas C.
; TITLE OF INVENTION: CDNA CLONING AND EXPRESSION OF HUMAN
; TITLE OF INVENTION: LIVER ESTROGEN SULFOTRANSFERASE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg & Woessner P.A.
; STREET: 3500 IDS Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/325,562
; FILING DATE: 18-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Muetting, Ann M.
; REGISTRATION NUMBER: 33,977
; REFERENCE/DOCKET NUMBER: 150.131US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-3061
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 294 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-325-562-2

Query Match      12.4%; Score 255; DB 1; Length 294;
Best Local Similarity 27.4%; Pred. No. 5.6e-16;
Matches 85; Conservative 51; Mismatches 120; Indels 54; Gaps 13;

QY 41 YKRYEIVSSLPNDYWGDTWRLYKGFQMGYLVPGIMAFEDN---FKARETDIIILTP 97
DB 8 YEKPEVHGIL-----MYKDF-----VKYWDNVFAQFARDLLVIATYP 46

QY 98 KAGTTWTKALTAIFAILTR-DVNHPSPSPHLLFFNPHSCVQNLEYLYMGRNTPDMLN 156
DB 47 KSGTTWSEIVYVNIYKGEDEVKCKED-----VIFN-----RIPFLECKENLMNGVKQLD 96

QY 157 E--SPRLFAGHIPYSLLPASVLKSGTKIINSRNKSTFVSFWKGNLI-----NPDKLLD 210
```



Db 97 EMNSPRIVKTHLPPELLPASFEKCKIIYLCRNAKDVAVSFYFFFLMVAGHPNGSFPPE 156  
Qy 211 LEKSYDIFASGISFCGPEWNPQAEFTNAASTNSNLLLSYEEMLEKPEVENVKLABFMGC 270  
Db 157 F---VEKFMQGVPG-SWKVHVSWEKGSFVLFYEDLKEDIRKEVILHFL-- 210  
Qy 271 GFTDDEKQGVDEIVKLCSPDNLAQOVNKGSSYNKIDNKH---FFRKGEVRDWAY 327  
Db 211 ---ERKPEBELVDRIIHTHSFOEMKNP-STNYTTLPLDEIMNOKLSFPMKRGITGDWKNH 266  
Qy 328 LTSEMIKLE 337  
Db 267 FTVALNEKFD 276

## RESULT 7

US-08-437-795-2  
; Sequence 2, Application US/08437795  
; Patent No. 5744355  
; GENERAL INFORMATION:  
; APPLICANT: Weinshilboum, Richard M.  
; APPLICANT: Aksoy, Ibrahim A.  
; APPLICANT: Wood, Thomas C.  
; TITLE OF INVENTION: CDNA CLONING AND EXPRESSION OF HUMAN  
; TITLE OF INVENTION: LIVER ESTROGEN SULFOTRANSFERASE  
; NUMBER OF SEQUENCES: 19  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Schwegman, Lundberg & Woessner P.A.  
; STREET: 3500 IDS Center  
; CITY: Minneapolis  
; STATE: MN  
; COUNTRY: USA  
; ZIP: 55402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/437,795  
; FILING DATE: 09-MAY-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Musting, Ann M.  
; REGISTRATION NUMBER: 33,977  
; REFERENCE/DOCKET NUMBER: 150.137US1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 612-339-0331  
; TELEFAX: 612-339-3061  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 294 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-437-795-2

Query Match 12.4%; Score 255; DB 1; Length 294;  
Best Local Similarity 27.4%; Pred. No. 5.6e-16;  
Matches 85; Conservative 51; Mismatches 120; Indels 54; Gaps 13;  
Qy 41 YKRYEIVSLPSNDYWGDTMBLYKGFQMGVYVGINAFEDN---FKARETDILITLTP 97  
Db 8 YKFEFVHGL-----MYKDF-----VKYMDNVEAFQARDDLVATYP 46  
Qy 98 KAGTTWKALTFAILTR-DVNHPSSTPHLPFPNPHSCVQNLLEYLYMGRENTMPDMLN 156  
Db 47 KSGTTWVSEIVMYKEGDEVCKED-----VIFN-----RIPFLECKENLWGVKQLD 96  
Qy 157 E--SPRLFAGHIPYSLLPASVLKSGTKIINISRNKSTFVSFWKFGNLI-----NPDKLID 210  
Db 97 EMNSPRIVKTHLPPELLPASFEKCKIIYLCRNAKDVAVSFYFFFLMVAGHPNGSFPPE 156

Qy 211 LEKSYDIFASGISFCGPEWNPQAEFTNAASTNSNLLLSYEEMLEKPEVENVKLABFMGC 270  
Db 157 F---VEKFMQGVPG-SWKVHVSWEKGSFVLFYEDLKEDIRKEVILHFL-- 210  
Qy 271 GFTDDEKQGVDEIVKLCSPDNLAQOVNKGSSYNKIDNKH---FFRKGEVRDWAY 327  
Db 211 ---ERKPEBELVDRIIHTHSFOEMKNP-STNYTTLPLDEIMNOKLSFPMKRGITGDWKNH 266  
Qy 328 LTSEMIKLE 337  
Db 267 FTVALNEKFD 276

## RESULT 8

US-09-609-816-10  
; Sequence 10, Application US/09609816  
; Patent No. 6436684  
; GENERAL INFORMATION:  
; APPLICANT: Woodage, Trevor  
; APPLICANT: Wei, Minh Hui  
; APPLICANT: Kodira, Chinappa  
; APPLICANT: Beasley, Ellen  
; APPLICANT: DiFrancesco, Valentina  
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING  
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN  
; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF  
; FILE REFERENCE: CL000669PCT  
; CURRENT APPLICATION NUMBER: US/09/609,816  
; CURRENT FILING DATE: 2000-07-03  
; PRIOR APPLICATION NUMBER: 60/192,408  
; PRIOR FILING DATE: 2000-03-27  
; PRIOR APPLICATION NUMBER: 60/212,725  
; PRIOR FILING DATE: 2000-06-20  
; PRIOR APPLICATION NUMBER: 09/609,816  
; PRIOR FILING DATE: 2000-07-03  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10  
; LENGTH: 283  
; TYPE: PRT  
; ORGANISM: HUMAN  
US-09-609-816-10

Query Match 12.0%; Score 246.5; DB 4; Length 283;  
Best Local Similarity 26.8%; Pred. No. 3.4e-15;  
Matches 71; Conservative 50; Mismatches 99; Indels 45; Gaps 9;  
Qy 83 NFKARETDILITLTPKAGTTWKALTFAI-----LTRDVNHPSSPTHLPLFFPHSCVQN 137  
Db 20 NFOAKPDDLLIATYAKAGTTWTQBIQNDSDVQCORANTYDRHPTI----- 69  
Qy 138 LEVLYMGRENTMP-----DLDMLE--SPRLFAGHIPYSLLPASVLKSGTKIINISRNK 190  
Db 70 -----EWTLPPPLNSGLDANKMPSPTLTKLTPVQMLPSPFWKENSQIIVARNAK 121  
Qy 191 STFVSFWKFGNLPNDKLL-----DLEKSDVIFASGISFCGPEWNPQAEFTNAASTNSNLL 246  
Db 122 DCLVSYIYFSRM---NKMPLDPGTLGEYIETFKAGKYLWG-SWYDHYKGVWDVYKDRHIL 177  
Qy 247 LLSYEMLEKPEVENVKLABFMGCFTDDBEKQGVDEIVKLCSPDNLAQOVNKGSSYN 306  
Db 178 YLFYEDMKEDPKREIKKIVKEL-----EKDISBEVLANKIHHHTSFDVWVKQNPWANYTLP 232  
Qy 307 NSKIDN--KHFFRKGEVRDWAYLT 329  
Db 233 SSIMDHSISPFMRKGMPGDWKNYFT 257

## RESULT 9

US-09-150-133-13  
; Sequence 13, Application US/09150133B  
; Patent No. 6060295

```

; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
; FILE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.504
; CURRENT APPLICATION NUMBER: US/09/150.133B
; CURRENT FILING DATE: 1998-09-09
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 5.1 (saved in ASCII format)
; SEQ ID NO 13
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: estrogen sulfotransferase
US-09-150-133-13

Query Match 11.9%; Score 245.5; DB 3; Length 295;
Best Local Similarity 25.5%; Pred. No. 4.5e-15;
Matches 82; Conservative 48; Mismatches 131; Indels 61; Gaps 12;

Qy 31 EQEKEEDSKMYKRYEIVSSLPNDYWGDTMRLYKGFQWQGYLVPGIMAFEDNFKARETD 90
Db 2 ETSMPYEVVGFGEFRGVLMDKRFKTKYWEDV-----EMFLARPDD 40

Qy 91 IILTLPKAGTTWTKALTFAILTR-DVNHPSPTPLFFNPHSCVQNLEYLYMGRENTM 149
Db 41 LVIAIYKSGTWTWSEVVMYIKGDEVECKEDA----IFN-----RIPYLECRNEDLI 90

Qy 150 PDLML--NESPLFAGHIPYSLLPASVLKSGTKIINISRNKSTFVSFWKFGNLI---- 203
Db 91 NGIKQLKEKESPRIVKTHLPKVLPAFWKCKMYLCRNKADKAVSYFFLLMITSYP 150

Qy 204 NPKSFSEF---VEKFMQGVPGY-SWYDHVKAWEKSKNSRVLFMFYEDMKEDIRREVVK 263
Db 151 NPKSFSEF---VEKFMQGVPGY-SWYDHVKAWEKSKNSRVLFMFYEDMKEDIRREVVK 206

Qy 264 LAEFMCGGFTDDEEKQIVDEIVKLCSPDNLKQOVNKGSSY-----NSKIDNKH 315
Db 207 LIEFL-----ERKPSAELVDRIIQTSTFQEMKNP-----STNYTWMPEEMNQKV--SPF 255

RESULT 11
US-09-374-493-13
; Sequence 13, Application US/09374493
; Patent No. 6204016
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.546
; CURRENT APPLICATION NUMBER: US/09/374,493
; CURRENT FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: 09/150,133
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 60/072,994
; EARLIER FILING DATE: 1998-01-29
; EARLIER APPLICATION NUMBER: PCT/US99/16750
; EARLIER FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 13
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: estrogen sulfotransferase
US-09-374-493-13

Query Match 11.9%; Score 245.5; DB 3; Length 295;
Best Local Similarity 25.5%; Pred. No. 4.5e-15;
Matches 82; Conservative 48; Mismatches 131; Indels 61; Gaps 12;

Qy 31 EQEKEEDSKMYKRYEIVSSLPNDYWGDTMRLYKGFQWQGYLVPGIMAFEDNFKARETD 90
Db 2 ETSMPYEVVGFGEFRGVLMDKRFKTKYWEDV-----EMFLARPDD 40

Qy 91 IILTLPKAGTTWTKALTFAILTR-DVNHPSPTPLFFNPHSCVQNLEYLYMGRENTM 149
Db 41 LVIAIYKSGTWTWSEVVMYIKGDEVECKEDA----IFN-----RIPYLECRNEDLI 90

Qy 150 PDLML--NESPLFAGHIPYSLLPASVLKSGTKIINISRNKSTFVSFWKFGNLI---- 203
Db 91 NGIKQLKEKESPRIVKTHLPKVLPAFWKCKMYLCRNKADKAVSYFFLLMITSYP 150

Qy 204 NPKSFSEF---VEKFMQGVPGY-SWYDHVKAWEKSKNSRVLFMFYEDMKEDIRREVVK 263
Db 151 NPKSFSEF---VEKFMQGVPGY-SWYDHVKAWEKSKNSRVLFMFYEDMKEDIRREVVK 206

Qy 264 LAEFMCGGFTDDEEKQIVDEIVKLCSPDNLKQOVNKGSSY-----NSKIDNKH 315
Db 207 LIEFL-----ERKPSAELVDRIIQTSTFQEMKNP-----STNYTWMPEEMNQKV--SPF 255

; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
; FILE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.495
; CURRENT APPLICATION NUMBER: US/09/150,141B
; CURRENT FILING DATE: 1998-09-09
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 5.1 (saved in ASCII format)
; SEQ ID NO 13
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: estrogen sulfotransferase
US-09-150-141-13

Query Match 11.9%; Score 245.5; DB 3; Length 295;
Best Local Similarity 25.5%; Pred. No. 4.5e-15;
Matches 82; Conservative 48; Mismatches 131; Indels 61; Gaps 12;

Qy 31 EQEKEEDSKMYKRYEIVSSLPNDYWGDTMRLYKGFQWQGYLVPGIMAFEDNFKARETD 90
Db 2 ETSMPYEVVGFGEFRGVLMDKRFKTKYWEDV-----EMFLARPDD 40

Qy 91 IILTLPKAGTTWTKALTFAILTR-DVNHPSPTPLFFNPHSCVQNLEYLYMGRENTM 149
Db 41 LVIAIYKSGTWTWSEVVMYIKGDEVECKEDA----IFN-----RIPYLECRNEDLI 90

Qy 150 PDLML--NESPLFAGHIPYSLLPASVLKSGTKIINISRNKSTFVSFWKFGNLI---- 203
Db 91 NGIKQLKEKESPRIVKTHLPKVLPAFWKCKMYLCRNKADKAVSYFFLLMITSYP 150

Qy 204 NPKSFSEF---VEKFMQGVPGY-SWYDHVKAWEKSKNSRVLFMFYEDMKEDIRREVVK 263
Db 151 NPKSFSEF---VEKFMQGVPGY-SWYDHVKAWEKSKNSRVLFMFYEDMKEDIRREVVK 206

Qy 264 LAEFMCGGFTDDEEKQIVDEIVKLCSPDNLKQOVNKGSSY-----NSKIDNKH 315
Db 207 LIEFL-----ERKPSAELVDRIIQTSTFQEMKNP-----STNYTWMPEEMNQKV--SPF 255

; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
; FILE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.495
; CURRENT APPLICATION NUMBER: US/09/150,141B
; CURRENT FILING DATE: 1998-09-09
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 5.1 (saved in ASCII format)
; SEQ ID NO 13
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: estrogen sulfotransferase
US-09-150-141-13

Query Match 11.9%; Score 245.5; DB 3; Length 295;
Best Local Similarity 25.5%; Pred. No. 4.5e-15;
Matches 82; Conservative 48; Mismatches 131; Indels 61; Gaps 12;

Qy 31 EQEKEEDSKMYKRYEIVSSLPNDYWGDTMRLYKGFQWQGYLVPGIMAFEDNFKARETD 90
Db 2 ETSMPYEVVGFGEFRGVLMDKRFKTKYWEDV-----EMFLARPDD 40

Qy 91 IILTLPKAGTTWTKALTFAILTR-DVNHPSPTPLFFNPHSCVQNLEYLYMGRENTM 149
Db 41 LVIAIYKSGTWTWSEVVMYIKGDEVECKEDA----IFN-----RIPYLECRNEDLI 90

Qy 150 PDLML--NESPLFAGHIPYSLLPASVLKSGTKIINISRNKSTFVSFWKFGNLI---- 203
Db 91 NGIKQLKEKESPRIVKTHLPKVLPAFWKCKMYLCRNKADKAVSYFFLLMITSYP 150

Qy 204 NPKSFSEF---VEKFMQGVPGY-SWYDHVKAWEKSKNSRVLFMFYEDMKEDIRREVVK 263
Db 151 NPKSFSEF---VEKFMQGVPGY-SWYDHVKAWEKSKNSRVLFMFYEDMKEDIRREVVK 206

Qy 264 LAEFMCGGFTDDEEKQIVDEIVKLCSPDNLKQOVNKGSSY-----NSKIDNKH 315
Db 207 LIEFL-----ERKPSAELVDRIIQTSTFQEMKNP-----STNYTWMPEEMNQKV--SPF 255
```

APPLICANT: The Board of Regents of Oklahoma  
TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF  
FILE REFERENCE: 5920 545  
CURRENT APPLICATION NUMBER: US/09/374,492  
CURRENT FILING DATE: 1999-08-13  
EARLIER APPLICATION NUMBER: 09/150,141  
EARLIER FILING DATE: 1998-09-09  
EARLIER APPLICATION NUMBER: 50/072,994

Db 2 ETSMEPEYVEGEFRGLMDKRFXYWEDV-----EMFLARPDD 40  
QY 91 IILITLPAAGTTWKALFAILLR-DVNHPSPTPLLFNPHSCVQNLVLYMGRENTM 149  
Db 41 LVIAIYPSKGTITWISSEVIMYKEDGVKCKEDA-----IFN-----RIPYLECRNEDLI 90  
QY 150 PDLML--NESPLPAGHIPYSLLPASVLKSGTKIINISNRKSTFTVSWFKGNLI----- 203  
Db 91 NGIKQKEKESPRIVKTHLPKVLPAFWEKNCCKMIYLCRNAKOVAVSYYYFLMLITSYP 150  
QY 204 NPKLLDLEKSDVIDFASGISFCGPENFQAETNAASTNSNLLLSYEEMLEKPEVNVK 263  
Db 151 NPKSPSEF---VEKMQGQVPG-SWYDHVKAWEKSKNSRVLFMFYEDMKEDIRREVVK 206  
QY 264 LAEFMCGCFDDEEKQGIIVDEIVKLCSEFDNLKQOVNKNNGSSY-----NSKIDNKH 315  
Db 207 LIEFL-----ERKPSAELVDRIQHTSFQEMKNP-----STNYTWPPEMMNQV--SPF 255  
QY 316 FRKGEVRDWANYLTSEMIXKLE 337  
Db 256 MKGIIGDWKNHPEALRERFD 277

RESULT 15

US-09-795-926-6  
; Sequence 6, Application US/09795926  
; Patent No. 6555669  
; GENERAL INFORMATION:  
; APPLICANT: Donoho, Gregory  
; APPLICANT: Hilbun, Erin  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Friedrich, Glenn  
; APPLICANT: Abuin, Alejandro  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sands, Arthur T.  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Wilganowski, Nathaniel L.  
; APPLICANT: Hu, Yi  
; APPLICANT: Kieke, James Alvin  
; APPLICANT: Potter, David George  
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
; FILE REFERENCE: LEX-0144-USA  
; CURRENT APPLICATION NUMBER: US/09/795,926  
; CURRENT FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: US 60/185,920  
; PRIOR FILING DATE: 2000-02-29  
; PRIOR APPLICATION NUMBER: US 60/186,558  
; PRIOR FILING DATE: 2000-03-02  
; PRIOR APPLICATION NUMBER: US 60/191,849  
; PRIOR FILING DATE: 2000-03-24  
; NUMBER OF SEQ ID NOS: 47  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 265  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-09-795-926-6

Query Match 11.6%; Score 239; DB 4; Length 265;  
Best Local Similarity 27.1%; Pred No 1.6e-14;  
Matches 79; Conservative 46; Mismatches 121; Indels 46; Gaps 12;  
QY 82 DNFKARETDIILTLPKAGTWT---TKALTFAILTRDVNHPSPHPLLFNPHSCVON 137  
Db 11 DTFEARHDDIVLASYPKCGSNWILHIVSELIYAVSKKYPFPV-----LECGDS 62  
QY 138 LEVLYMGRENTMPDMLNESPRLFAGHIPYSLLPASVLKSGTKIINISNRKSTFTVSEW 197  
Db 63 EKY---QRMKGFP-----SPRILATHLYDKLPGSIFENKAKILVIFRNPKDTAVSFL 112  
QY 198 KFGNLIINPKLLDLKESVDIFASGISFCGPENFQAET---NAASTNSNLLLSYEML 254

Db 113 HFHNDV-PD--IPSYGSWDEFFRQPMKQVSGRYPDFDFAINWKNKHLDCGNVAFILYEDLK 169  
QY 255 EKEVENVKKLABFMCGCGFTDDEEKQGIIVDEIVKLCSEFDNLKQOVNKNNGSSYNSKIDNKH 314  
Db 170 ENLAAGIKQIAEFLGPFLLTGEQ-----IQTISVQSTFQAMRAKSQDTHGAV-----GPF 218  
QY 315 FFRKGEVRDWANYLTSEMIXKLETAGKINSEKHLISNKNL-TNFWLNNYCE 365  
Db 219 LFRKGEVGDWKN-LFSEIQNQ-----EMDEKFKBCLAGTSLGAKLKYESYCO 264

Search completed: March 20, 2004, 01:39:19  
Job time : 18 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2004, 01:37:48 ; Search time 34 Seconds  
(without alignments)  
2977.986 Million cell updates/sec

Title: US-09-854-122-16

Perfect score: 2059  
Sequence: 1 TRGTGIAVACLPLIMAGIL.....CLLSYSFVINNFILLKXXXK 391

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1049977 seqs, 258955339 residues

Total number of hits satisfying chosen parameters: 1049977

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US05\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US05\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2059	100.0	391	9	US-09-854-122-16
2	1750	85.0	331	9	US-09-854-122-17
3	589	28.6	344	12	US-10-424-599-201713
4	527.5	25.6	302	9	US-09-854-122-19
5	525.5	25.5	365	12	US-10-425-114-63828
6	525.5	25.5	365	12	US-10-425-114-65063
7	521.5	25.3	338	15	US-10-259-194A-4
8	519.5	25.2	348	15	US-10-259-194A-380
9	510.5	24.8	320	9	US-09-854-122-20
10	501.5	24.4	328	14	US-10-259-165-114
11	501.5	24.4	328	14	US-10-259-165-444
12	500.5	24.3	324	9	US-09-854-122-18
13	483	23.5	688	14	US-10-195-144-79
14	483	23.5	688	15	US-10-345-072-79
15	479	23.3	343	15	US-10-259-194A-330

16	446.5	21.7	346	14	US-10-259-165-332	Sequence 332, App
17	444.5	21.6	251	12	US-10-424-599-225084	Sequence 225084,
18	441.5	21.4	486	15	US-10-259-194A-132	Sequence 132, App
19	363.5	17.7	349	15	US-10-259-194A-16	Sequence 16, Appl
20	309	15.0	269	12	US-10-072-012-855	Sequence 855, App
21	309	15.0	269	12	US-10-072-012-867	Sequence 867, App
22	304	14.8	302	12	US-10-072-012-586	Sequence 586, App
23	288	13.0	304	13	US-10-199-330-6	Sequence 6, Appl
24	288	13.0	304	14	US-10-199-334-6	Sequence 6, Appl
25	268	13.0	304	14	US-10-199-329-8	Sequence 585, App
26	266.5	12.9	307	12	US-10-072-012-585	Sequence 206, App
27	265.5	12.9	295	12	US-10-072-012-206	Sequence 582, App
28	264	12.8	304	12	US-10-072-012-582	Sequence 4, Appl
29	264	12.8	304	13	US-10-199-330-4	Sequence 4, Appl
30	264	12.8	304	14	US-10-199-334-4	Sequence 4, Appl
31	264	12.8	304	14	US-10-199-329-4	Sequence 5, Appl
32	260	12.6	304	13	US-10-199-330-5	Sequence 5, Appl
33	260	12.6	304	14	US-10-199-334-5	Sequence 5, Appl
34	260	12.6	304	14	US-10-199-329-5	Sequence 5, Appl
35	255.5	12.4	283	9	US-09-898-570-40	Sequence 40, Appl
36	255.5	12.4	283	10	US-09-839-446-40	Sequence 40, Appl
37	255.5	12.4	283	13	US-10-199-330-8	Sequence 8, Appl
38	255.5	12.4	283	13	US-10-199-330-9	Sequence 9, Appl
39	255.5	12.4	283	14	US-10-199-334-8	Sequence 8, Appl
40	255.5	12.4	283	14	US-10-199-334-9	Sequence 9, Appl
41	255.5	12.4	283	14	US-10-199-329-9	Sequence 8, Appl
42	255.5	12.4	283	14	US-10-199-329-9	Sequence 9, Appl
43	255.5	12.4	304	12	US-10-072-012-583	Sequence 583, App
44	253.5	12.3	305	9	US-09-898-570-26	Sequence 26, Appl
45	253.5	12.3	305	10	US-09-839-446-26	Sequence 26, Appl

#### ALIGNMENTS

##### RESULT 1

US-09-854-122-16  
; Sequence 16, Application US/09854122  
; Patent No. US20020016980A1  
; GENERAL INFORMATION:  
; APPLICANT: SMITH, ROBERT  
; TITLE OF INVENTION: TRANSGENIC PLANTS INCORPORATING TRAIT OF ZOSTERA MARINA  
; FILE REFERENCE: PHA-007.01  
; CURRENT APPLICATION NUMBER: US/09/854,122  
; PRIOR FILING DATE: 2001-09-10  
; PRIOR APPLICATION NUMBER: 60/202,529  
; PRIOR FILING DATE: 2000-05-10  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 16  
; LENGTH: 391  
; TYPE: PRT  
; ORGANISM: Zostera marina  
; US-09-854-122-16

Query Match	100.0%	Score 2059;	DB 9;	Length 391;
Best Local Similarity	100.0%;	Pred. No. 4.5e-185;		
Matches 391;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	TRGTGIAVACLPLIMAGILALEKFCGSKNEQKEEDSKMYKRYREIVSSLPNDYWGDT	60	
Db	1	TRGTGIAVACLPLIMAGILALEKFCGSKNEQKEEDSKMYKRYREIVSSLPNDYWGDT	60	
Qy	61	MELYKGFQWQGYLVPFGIMAFEDNFKAETDIIITLTPKAGTTWTKALTFAILTRDVNHP	120	
Db	61	MELYKGFQWQGYLVPFGIMAFEDNFKAETDIIITLTPKAGTTWTKALTFAILTRDVNHP	120	
Qy	121	SPTHPLFFNPHSCVQNLEYLYMGRENTMPDMLNESPRLFAGHIPYSLLPASVLKSGT	180	
Db	121	SPTHPLFFNPHSCVQNLEYLYMGRENTMPDMLNESPRLFAGHIPYSLLPASVLKSGT	180	
Qy	181	KIINISRNKSTFTFSVFWKFGNLIINPDKLLDLKSVVDIFASGISFCGPEWNFQAEFTNAAS	240	

Db 181 KIINISRNKSTFVSFWKPGNLINPKLLDLEKSDVDFASGISFCGPEWFOAFTNAAS 240  
QY 241 TNSNLLLSYEEMLEKPVENVKLAEFMCGGFTDDEEKQGVDEIVKLCFDFNLKQOVN 300  
Db 241 TNSNLLLSYEEMLEKPVENVKLAEFMCGGFTDDEEKQGVDEIVKLCFDFNLKQOVN 300  
QY 301 KNGSSVNSKIDNKHFRKGEVRDWNLYTSEMICKLETAGKINSEKHLNKNLTNFKL 360  
Db 301 KNGSSVNSKIDNKHFRKGEVRDWNLYTSEMICKLETAGKINSEKHLNKNLTNFKL 360  
QY 361 NNYCKSNISLCLLSYSFVNNFILLKKKK 391  
Db 361 NNYCKSNISLCLLSYSFVNNFILLKKKK 391

RESULT 2  
US-09-854-122-17  
; Sequence 17, Application US/09854122  
; Patent No. US20020016980A1  
; GENERAL INFORMATION:  
; APPLICANT: SMITH, ROBERT  
; TITLE OF INVENTION: TRANSGENIC PLANTS INCORPORATING TRAIT OF ZOSTERA MARINA  
; FILE REFERENCE: PHA-007.01  
; CURRENT APPLICATION NUMBER: US/09/854,122  
; PRIOR FILING DATE: 2001-09-10  
; PRIOR FILING DATE: 2000-05-10  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 17  
; LENGTH: 331  
; TYPE: PRT  
; ORGANISM: Zostera marina  
US-09-854-122-17

Query Match 85.0%; Score 1750; DB 9; Length 331;  
Best Local Similarity 100.0%; Pred. No. 4.4e-156;  
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 MAGILALEKCFGSKNEQKEEDSKMYKRYRIVASLSPSNDYWGDTMRLYKGFQMGYLVP 75  
Db 1 MAGILALEKCFGSKNEQKEEDSKMYKRYRIVASLSPSNDYWGDTMRLYKGFQMGYLVP 60

QY 76 GIMAFEDNFKARETDIITLTPKAGTTWTKALTFAITRDVNHPSPTPLFPNPHSCV 135  
Db 61 GIMAFEDNFKARETDIITLTPKAGTTWTKALTFAITRDVNHPSPTPLFPNPHSCV 120

QY 136 ONLEYLYMGRENTMPDLMLNESPRLFAGHIPYSLIPASVLKSGTKIINISRNKSTFVS 195  
Db 121 ONLEYLYMGRENTMPDLMLNESPRLFAGHIPYSLIPASVLKSGTKIINISRNKSTFVS 180

QY 196 FWKFGNLINPKLLDLEKSDVDFASGISFCGPEWFOAFTNAASTNSNLLLSYEEMLE 255  
Db 181 FWKFGNLINPKLLDLEKSDVDFASGISFCGPEWFOAFTNAASTNSNLLLSYEEMLE 240

QY 256 KPVENVKLAEFMCGGFTDDEEKQGVDEIVKLCFDFNLKQOVNKGSSVNSKIDNKH 315  
Db 241 KPVENVKLAEFMCGGFTDDEEKQGVDEIVKLCFDFNLKQOVNKGSSVNSKIDNKH 300

QY 316 FRKGEVRDWNLYTSEMICKLETAGKINSE 346  
Db 301 FRKGEVRDWNLYTSEMICKLETAGKINSE 331

RESULT 3  
US-10-424-599-201713  
; Sequence 201713, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 201713  
; LENGTH: 344  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_24172C.1.pap  
US-10-424-599-201713

Query Match 28.6%; Score 589; DB 12; Length 344;  
Best Local Similarity 38.6%; Pred. No. 1e-46;  
Matches 125; Conservative 69; Mismatches 116; Indels 14; Gaps 8;

QY 25 CFGSKNEQKE-----BDSKMYKRYRIVASLSPSNDYWGDT-MRLYKGFQMGYLVP 78  
Db 8 CFREENSEKGEETIEEDKLSQECKELILSLPRGWRTRYIYLFQGFQWCPLEIQAI 67

QY 79 AFEDNFKARETDIITLTPKAGTTWTKALTFAITRDVNH--PSSPTPLFPNPHSCV 136  
Db 68 TFQHFQAKSDVIVATIPKSGITLWLTALFAIVNRHTSITTSMSHPLTNSPHELVP 127

QY 137 NLEY-LYMGRENTMPDLMLNESPRLFAGHIPYSLIPASVLKSGTKIINISRNKSTFVS 195  
Db 128 FIEYTVGNAPSHVPLNSNMT-PRLFQTHIPHALAKSIKESNRRIIVCRNPLDTFVS 186

QY 196 FWKFGNLINPKLLDLEKSDVDFASGISFCGPEWFOAFTNAASTNSNLLLSYE 252  
Db 187 TWIFLNKIPHELFELGAFKPKYKGIIGFTWQMLGYWKESIAKSPKVLFLKYED 246

QY 253 MLEKPEVNVKLAEFMCGGFTDDEEKQGVDEIVKLCSPDNKXQVKNKGSSVNSKIDN 312  
Db 247 LKQVNVFHVKRIAEFLGCPFTSBEEDGDTIESIILCSPEKMELEANKSG-TFARNFER 305

QY 313 KHFRKGEVRDWNLYTSEMICKL 336  
Db 306 KILFRKABMGDWNLYSPFEMGEKL 329

RESULT 4  
US-09-854-122-19  
; Sequence 19, Application US/09854122  
; Patent No. US20020016980A1  
; GENERAL INFORMATION:  
; APPLICANT: ALBERTE, RANDALL S.  
; APPLICANT: SMITH, ROBERT  
; TITLE OF INVENTION: TRANSGENIC PLANTS INCORPORATING TRAIT OF ZOSTERA MARINA  
; FILE REFERENCE: PHA-007.01  
; CURRENT APPLICATION NUMBER: US/09/854,122  
; CURRENT FILING DATE: 2001-09-10  
; PRIOR APPLICATION NUMBER: 60/202,529  
; PRIOR FILING DATE: 2000-05-10  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 19  
; LENGTH: 302  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
US-09-854-122-19

Query Match 25.6%; Score 527.5; DB 9; Length 302;  
Best Local Similarity 41.1%; Pred. No. 5.3e-44;  
Matches 122; Conservative 47; Mismatches 107; Indels 21; Gaps 9;

QY 37 DSKMYKRYRIVASLSPSNDYWGDTMRLYKGFQMGYLVP 95  
Db 13 DEDLTQETRALISSLPKEKGLWVSEIVFQGLWHTQAILQILQCKRFEAKDSIILVT 72

```
QY 96 LPKAGTTWKALTFALTRDVNH-----PSSPTPLLPFNPHSCVONLEYLYMGRENTMPD 151
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 73 NPKSGTTWLKALVALLNR---HKRPVSSSGNHPLLVNTNPHLLVPFLEGVYV---ESP 125
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 152 LDMLN-ESPRLPAGHFPYSLPASVLKSGTKIINISRNKSTFVYFWKGNLNPDKLLD 210
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 126 FDFSLPSFLNTHISLSLPSVSKSSCKVICCRNPKDMFVSLWHFGKLAPEETAD 185
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 211 --LEKSVDFIFAGISFCGPEWFOAEFTNAASTNSN-LLLSYEEMLEKPVYKLAEF 267
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 186 YPIEKAVEAFCEGKFIGGFWDHILEYVYASRENPKVLFVYBELKKQTEVEMKRIAEF 245
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 268 MCGGTDEEEKQGVDEIVKLCSPDNLKNQKNGSSYNKIDNKHFFRKGEVRDW 324
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 246 LECGFIEBEE---VREIVKLCSPESLNLEVNKEGLPNG-IETKFFFRKGEGGW 297
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 5
US-10-425-114-63828
; Sequence 63828, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 63828
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3353-009-H8_FLI.pep
US-10-425-114-63828
Query Match 25.5%; Score 525.5; DB 12; Length 365;
Best Local Similarity 35.2%; Pred. No. 1.1e-40;
Matches 116; Conservative 65; Mismatches 114; Indels 35; Gaps 11;
QY 30 NEOKEED--SKMYKRYRIVSVSLPSNDYWGDT-----MRLYKGFWMQGYL---VPGI 77
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 31 NQAEASPLASPHSNIAIISLPL-----ETRCPPFLRYANFWVPEVILKADLPGI 85
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 78 MAFEDNFKARETDIILTLPKAGTTWKALTFALTRDVNHPSSTHPLLPFNPHSCVON 137
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 86 ---HSCFKPRPTDVFVASFPKSGTTWLKALAFATLKRSTHPLDGDHPLRRCNPHDCVRF 142
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 138 LEVLYMGRENTMPDLDMLNESPLPAGHFPYSLPASVL---KSGTKIINISRNKSTF 193
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 143 LDANFNQCKDEAL-----PSRVLATHLPYSLPGSITGDRRSGCRIVYVCRPKDAL 198
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 194 VSFWKPF---GNLINPD-KLLDLEKSVDFIFAGISFCGPEWFOAEF-TNAASTNSNLL 247
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 199 VSAWLFRKAASALGADARSFTIQEALFLCDGRCMCGPQWEHVLQYWEESVRRPDRVLF 258
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 248 LSYEEMLEKPVENYKLAEMFGGFTDDEEKQGVDEIVKLCFDFNLKXQVKNKSSYN 307
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 259 LRYEEMLIDPEAHVRKLAKFMGCGFSEBEEHGVSAIVELCSLKGWRDNEVNRNGSNML 318
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 308 SKIDNKHFFRKGEVRDWANYLTSEMICKLE 337
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 319 G-VKNESYFRKGVAGDSNEMTDPMAQRLD 347
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 6
US-10-425-114-65063
; Sequence 65063, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 65063
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB4740-009-D3_FLI.pep
US-10-425-114-65063
Query Match 25.5%; Score 525.5; DB 12; Length 365;
Best Local Similarity 35.2%; Pred. No. 1.1e-40;
Matches 116; Conservative 65; Mismatches 114; Indels 35; Gaps 11;
QY 30 NEOKEED--SKMYKRYRIVSVSLPSNDYWGDT-----MRLYKGFWMQGYL---VPGI 77
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 31 NQAEASPLASPHSNIAIISLPL-----ETRCPPFLRYANFWVPEVILKADLPGI 85
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 78 MAFEDNFKARETDIILTLPKAGTTWKALTFALTRDVNHPSSTHPLLPFNPHSCVON 137
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 86 ---HSCFKPRPTDVFVASFPKSGTTWLKALAFATLKRSTHPLDGDHPLRRCNPHDCVRF 142
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 138 LEVLYMGRENTMPDLDMLNESPLPAGHFPYSLPASVL---KSGTKIINISRNKSTF 193
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 143 LDANFNQCKDEAL-----PSRVLATHLPYSLPGSITGDRRSGCRIVYVCRPKDAL 198
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 194 VSFWKPF---GNLINPD-KLLDLEKSVDFIFAGISFCGPEWFOAEF-TNAASTNSNLL 247
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 199 VSAWLFRKAASALGADARSFTIQEALFLCDGRCMCGPQWEHVLQYWEESVRRPDRVLF 258
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 248 LSYEEMLEKPVENYKLAEMFGGFTDDEEKQGVDEIVKLCFDFNLKXQVKNKSSYN 307
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 259 LRYEEMLIDPEAHVRKLAKFMGCGFSEBEEHGVSAIVELCSLKGWRDNEVNRNGSNML 318
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 308 SKIDNKHFFRKGEVRDWANYLTSEMICKLE 337
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 319 G-VKNESYFRKGVAGDSNEMTDPMAQRLD 347
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 7
US-10-259-194A-4
; Sequence 4, Application US/10259194A
; Publication No. US2004010815A1
; GENERAL INFORMATION:
; APPLICANT: Lange, Markus B.
; APPLICANT: Grasseman, Majid
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kreps, Joel
; APPLICANT: Moughamer, Todd
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF PLANT GENES
; FILE REFERENCE: 70029-NP
; CURRENT APPLICATION NUMBER: US/10/259,194A
```

263 KLAFFMCGFTDDEEKQIVDEIVKLCSPDNLKNQVKNKSSYNSK---IDNKHFFRK 318

Y 263



Db 230 RLAETFGVFTTBEKEGVIEIILKCSFENLSLEVNSKSG---NSKGFLEIENRLYPRK 286  
Qy 319 GEVRDWANYLTSEMICKLE 337  
Db 287 AKDGDKNYFTDTEMKID 305

## RESULT 10

US-10-259-165-114  
; Sequence 114, Application US/10259165  
; Publication No. US2003013588A1

## GENERAL INFORMATION:

; APPLICANT: Zhu, Tong  
; APPLICANT: Wang, Xun  
; APPLICANT: Chang, Hur-song  
; APPLICANT: Briggs, Steven P.  
; APPLICANT: Cooper, Bret  
; APPLICANT: Glazebrook, Jane  
; APPLICANT: Goff, Stephen A.  
; APPLICANT: Katagiri, Fumiaki  
; APPLICANT: Kreps, Joel  
; APPLICANT: Moughamer, Todd  
; APPLICANT: Provart, Nicholas  
; APPLICANT: Rickel, Darrell

; TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE SILENCING

; FILE REFERENCE: 70030-NP  
; CURRENT APPLICATION NUMBER: US/10/259,165  
; CURRENT FILING DATE: 2002-09-26  
; PRIOR APPLICATION NUMBER: US 60/370,620  
; PRIOR FILING DATE: 2002-04-04  
; PRIOR APPLICATION NUMBER: US 60/368,327  
; PRIOR FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: US 60/325,277  
; PRIOR FILING DATE: 2001-09-26

; NUMBER OF SEQ ID NOS: 782  
; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta

; SEQ ID NO 114

; LENGTH: 328

; TYPE: PRT

; ORGANISM: Oryza sativa

US-10-259-165-114

Query Match 24.4%; Score 501.5; DB 14; Length 328;  
Best Local Similarity 38.3%; Pred. No. 1.7e-38;  
Matches 119; Conservative 39; Mismatches 106; Indels 47; Gaps 8;

Qy 48 VSSLPNDYMGD-TMRLYKGFWMGVLPVGINAFEDNFKARETDIILTLPKAGTTWTKA 106  
Db 25 ISSLPLETRCAPPFLRQHGFWLPETFLPGLEAARARFPRPSDVLLASFPSKSGTTWLKA 84  
Qy 107 LTFALTRDYNHPSSTHPLLFNPHSCVQNLEYLYMGRENTMPDLDMLNESPRLFAGHI 166  
Db 85 LAFATLNRAAYPSGEGHPLRRRGPHDCVQFLESALVSDDMFASL----PSPRLLSHTL 140  
Qy 167 PYSLLPASVL--KSGTKIINISNRKSTFVSFWKF-----GN---LINPKD 207  
Db 141 PYSLLPEGKADSSGCRIDYICRDPKDVLSVWMLFTKALGTODGPTNGNKPMLSN--- 197  
Qy 208 LLDLEKSVDFASGISFCGPEWNFQAEF-TNAASTNSNLLLSYEEMLEKPVENVKLA 266  
Db 198 -----GPNRHVLEYWAESKRPRQKVLFLRYEBMTRETTTSNVKLA 239  
Qy 267 FMCGFTDDBEKGQIVDEIVKLCSPFNKQVKNKNGSSYNKIDNKHFFRKGVRDWN 326  
Db 240 FMGCPFSGEEADGVDPDAIVGLCSFDHLRSLEVNRNGAN-DFNKNDSFYRKGVDWAN 298  
Qy 327 YLTSEMIKLE 337  
Db 299 YLSPENAAQLD 309

## RESULT 11

US-10-259-165-444  
; Sequence 444, Application US/10259165  
; Publication No. US2003013588A1

## GENERAL INFORMATION:

; APPLICANT: Zhu, Tong  
; APPLICANT: Wang, Xun  
; APPLICANT: Chang, Hur-song  
; APPLICANT: Briggs, Steven P.  
; APPLICANT: Cooper, Bret  
; APPLICANT: Glazebrook, Jane  
; APPLICANT: Goff, Stephen A.  
; APPLICANT: Katagiri, Fumiaki  
; APPLICANT: Kreps, Joel  
; APPLICANT: Moughamer, Todd  
; APPLICANT: Provart, Nicholas  
; APPLICANT: Rickel, Darrell

; TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE SILENCING

; FILE REFERENCE: 70030-NP  
; CURRENT APPLICATION NUMBER: US/10/259,165  
; CURRENT FILING DATE: 2002-09-26  
; PRIOR APPLICATION NUMBER: US 60/370,620  
; PRIOR FILING DATE: 2002-04-04  
; PRIOR APPLICATION NUMBER: US 60/368,327  
; PRIOR FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: US 60/325,277  
; PRIOR FILING DATE: 2001-09-26

; NUMBER OF SEQ ID NOS: 782  
; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta

; SEQ ID NO 444

; LENGTH: 328

; TYPE: PRT

; ORGANISM: Oryza sativa

US-10-259-165-444

Query Match 24.4%; Score 501.5; DB 14; Length 328;  
Best Local Similarity 38.3%; Pred. No. 1.7e-38;  
Matches 119; Conservative 39; Mismatches 106; Indels 47; Gaps 8;

Qy 48 VSSLPNDYMGD-TMRLYKGFWMGVLPVGINAFEDNFKARETDIILTLPKAGTTWTKA 106  
Db 25 ISSLPLETRCAPPFLRQHGFWLPETFLPGLEAARARFPRPSDVLLASFPSKSGTTWLKA 84  
Qy 107 LTFALTRDYNHPSSTHPLLFNPHSCVQNLEYLYMGRENTMPDLDMLNESPRLFAGHI 166  
Db 85 LAFATLNRAAYPSGEGHPLRRRGPHDCVQFLESALVSDDMFASL----PSPRLLSHTL 140  
Qy 167 PYSLLPASVL--KSGTKIINISNRKSTFVSFWKF-----GN---LINPKD 207  
Db 141 PYSLLPEGKADSSGCRIDYICRDPKDVLSVWMLFTKALGTODGPTNGNKPMLSN--- 197  
Qy 208 LLDLEKSVDFASGISFCGPEWNFQAEF-TNAASTNSNLLLSYEEMLEKPVENVKLA 266  
Db 198 -----GPNRHVLEYWAESKRPRQKVLFLRYEBMTRETTTSNVKLA 239  
Qy 267 FMCGFTDDBEKGQIVDEIVKLCSPFNKQVKNKNGSSYNKIDNKHFFRKGVRDWN 326  
Db 240 FMGCPFSGEEADGVDPDAIVGLCSFDHLRSLEVNRNGAN-DFNKNDSFYRKGVDWAN 298  
Qy 327 YLTSEMIKLE 337  
Db 299 YLSPENAAQLD 309

## RESULT 12

US-09-854-122-18

; Sequence 18, Application US/09854122  
; Patent No. US20020016980A1

## GENERAL INFORMATION:

; APPLICANT: SMITH, ROBERT  
; APPLICANT: ALBERTE, RANDALL S.

; TITLE OF INVENTION: TRANSGENIC PLANTS INCORPORATING TRAIT OF ZOSTERA MARINA  
; FILE REFERENCE: PHA-007.01  
; CURRENT APPLICATION NUMBER: US/09/854,122

```
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/202,529
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 324
; TYPE: PRT
; ORGANISM: Brassica napus
US-09-854-122-18

Query Match      24.3%; Score 500.5; DB 9; Length 324;
Best Local Similarity 38.0%; Pred. No. 2.1e-38;
Matches 117; Conservative 58; Mismatches 114; Indels 19; Gaps 10;

QY 37 DSKMKRYREIVSSPSNDY-GDTRLYKGFQWQVGLVPGIMAFEDNFKARETDIILTT 95
DB 13 DENLTQKTKYDLISSPSEKGLVQMTQFQGRWTOALLOGLTCQKHPAKOSDIILVT 72
QY 96 LPKAGTTWKALTFAILTRDVNHPSSPHLLFFNPHSCVQN-LEYLYMGRE-NTMPDLD 153
DB 73 NPKSGTTWKLKALVFAL----INRHKFPVYSVIL---SCYQSALLVPFLGRSLRSPDFD 125
QY 154 MLN-ESPLFAGHIPYSLLPASVLKSGTKIINISRNKSTFVGFQKGNLINPKLLD-- 210
DB 126 FSQLSPLMNTSHLSLPSVKSSCKIYCCNPKDMFVSLWHPGKLAPEETADYP 185
QY 211 LKSVVDIFASGISFCGPEWNNFQAEFTNAASTNSN-LILLSYEEMLEKPVENVKLAEFWG 269
DB 186 IEKAVEAPCOQKFIGGPPFWHVLVYVYASLENPNKVLFSVEEPKKTGTETIKRIAEFLG 245
QY 270 COFTDDEKQKQIVDEIVKLCFSDNLKQOVNKGSSVNSKIDNKHFRKGEVDRWANYLT 329
DB 246 COLVGEER----VRAIVKLCFSLSLSEVNRG-KLPDSGNETFAFRKGEVGGHRTLT 300
QY 330 SEMIKKLE 337
DB 301 ESIAEVID 308

RESULT 13
US-10-195-144-79
; Sequence 79, Application US/10195144
; Publication No. US20030126646A1
; GENERAL INFORMATION:
; APPLICANT: BROWN, GREGORY G.
; APPLICANT: FORMANOVA, NATASA
; APPLICANT: DENDY, CHARLES
; APPLICANT: LANDRY, BENOIT S.
; APPLICANT: CHEUNG, WING
; APPLICANT: JIN, HUA
; TITLE OF INVENTION: NUCLEAR FERTILITY RESTORER GENES AND METHODS OF USE IN
; FILE REFERENCE: 16313-0136
; CURRENT APPLICATION NUMBER: US/10/195,144
; PRIOR FILING DATE: 2002-10-01
; PRIOR FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: 60/305,026
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: 60/305,363
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/308,736
; NUMBER OF SEQ ID NOS: 128
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 79
; LENGTH: 688
; TYPE: PRT
; ORGANISM: Raphanus sativum
US-10-195-144-79

Query Match      23.5%; Score 483; DB 14; Length 688;
Best Local Similarity 37.8%; Pred. No. 2.9e-36;
Matches 107; Conservative 55; Mismatches 107; Indels 14; Gaps 7;

QY 56 YWGTMLYKGFQWQVGLVPGIMAFEDNFKARETDIILTTLPKAGTTWKALTFAILTRD 115
DB 78 FFGGKLYFSILAGNYIFPVLRGQ----PODVTIIASYPKSGTTWKLKALTVALLE 133
QY 116 VNHPSSTPHLLFPNPHSCVQNLEYLYMGRENTMPDLDMLNESPRLFAGHIPYSLLPASV 175
DB 134 KQHSSD--HPLLYNPHGIIPFLE-IDVYHSSSNLAKFSAPPLFSTHMLPHIHEAL 190
QY 176 LKSGTKIINISRNKSTFVSVFWKFG--NLINPKLLDLEKSVDFIPASGISFCGPEWNNFQ 232
DB 191 KHSCKIIVYCRNVKOTLISCNFYSCAIYKIEPTRRV-LESFNEFCGDTNFGFPWDHL 249
QY 233 AEFTNAASTN-SNLLLSYEEMLEKPVENVKLAEFWGCGFTDDEKQKQIVDEIVKLCF 291
DB 250 LSYWRGSLDEPKHVLFWYEEEMKAEPRDQIKRLADFLGCGFTKQBEDSSVSGILDLSL 309
QY 292 DNLKQOVNKGSSVNSKIDNKHFRKGEVDRWANYLTSEMIX 334
DB 310 RNLSSLEANKTGTTINN--VEHKFFFRKGEVGGSKNYLTSEMEX 350

RESULT 14
US-10-345-072-79
; Sequence 79, Application US/10345072
; Publication No. US20030237112A1
; GENERAL INFORMATION:
; APPLICANT: BROWN, GREGORY G.
; APPLICANT: FORMANOVA, NATASA
; APPLICANT: DENDY, CHARLES
; APPLICANT: LANDRY, BENOIT S.
; APPLICANT: CHEUNG, WING
; APPLICANT: JIN, HUA
; APPLICANT: LAI, FANG MING
; APPLICANT: LEFOREST, MARTIN
; TITLE OF INVENTION: NUCLEAR FERTILITY RESTORER GENES AND METHODS OF USE IN
; FILE REFERENCE: 16313-0210
; CURRENT APPLICATION NUMBER: US/10/345,072
; PRIOR FILING DATE: 2003-01-16
; PRIOR APPLICATION NUMBER: PCT/US02/22217
; PRIOR FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 60/305,026
; PRIOR FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: 60/305,363
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: 60/308,736
; PRIOR FILING DATE: 2001-07-30
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 79
; LENGTH: 688
; TYPE: PRT
; ORGANISM: Raphanus sativum
US-10-345-072-79

Query Match      23.5%; Score 483; DB 15; Length 688;
Best Local Similarity 37.8%; Pred. No. 2.9e-36;
Matches 107; Conservative 55; Mismatches 107; Indels 14; Gaps 7;

QY 56 YWGTMLYKGFQWQVGLVPGIMAFEDNFKARETDIILTTLPKAGTTWKALTFAILTRD 115
DB 78 FFGGKLYFSILAGNYIFPVLRGQ----PODVTIIASYPKSGTTWKLKALTVALLE 133
QY 116 VNHPSSTPHLLFPNPHSCVQNLEYLYMGRENTMPDLDMLNESPRLFAGHIPYSLLPASV 175
DB 134 KQHSSD--HPLLYNPHGIIPFLE-IDVYHSSSNLAKFSAPPLFSTHMLPHIHEAL 190
QY 176 LKSGTKIINISRNKSTFVSVFWKFG--NLINPKLLDLEKSVDFIPASGISFCGPEWNNFQ 232
DB 191 KHSCKIIVYCRNVKOTLISCNFYSCAIYKIEPTRRV-LESFNEFCGDTNFGFPWDHL 249
QY 233 AEFTNAASTN-SNLLLSYEEMLEKPVENVKLAEFWGCGFTDDEKQKQIVDEIVKLCF 291
```

Job time : 35 secs

Db 250 LSYWGSLEDKXVLFMYEENKAEPRQIKRLADFLCPTKQEESSGSDGILDCSL 309  
QY 292 DNLKNQVKNQSSVNSKDNKHPKRGKGEVDMANYLTSEMIX 334  
Db 310 RNLSSLEANKTGTINN--VEHKFFPKRGKGEVDSKNYLTSEMEX 350

RESULT 15  
US-10-259-194A-330  
; Sequence 330, Application US/10259194A  
; Publication No. US20040010815A1  
; GENERAL INFORMATION:  
; APPLICANT: Lange, Markus B.  
; APPLICANT: Ghassseman, Majid  
; APPLICANT: Briggs, Steven P.  
; APPLICANT: Cooper, Bret  
; APPLICANT: Glazebrook, Jane  
; APPLICANT: Goff, Stephen A.  
; APPLICANT: Katagiri, Fumiaki  
; APPLICANT: Kreps, Joel  
; APPLICANT: Moughamer, Todd  
; APPLICANT: Provart, Nicholas  
; APPLICANT: Ricks, Darrell  
; APPLICANT: Zhu, Tong  
; TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF PLANT GENES  
; FILE REFERENCE: 70029-NP  
; CURRENT APPLICATION NUMBER: US/10/259,194A  
; CURRENT FILING DATE: 2003-01-07  
; PRIOR APPLICATION NUMBER: US 60/325,277  
; PRIOR FILING DATE: 2001-09-26  
; PRIOR APPLICATION NUMBER: US 60/370,743  
; PRIOR FILING DATE: 2002-04-04  
; PRIOR APPLICATION NUMBER: US 60/370,620  
; PRIOR FILING DATE: 2002-04-04  
; NUMBER OF SEQ ID NOS: 662  
; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta  
; SEQ ID NO 330  
; LENGTH: 343  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-259-194A-330

Query Match 23.3%; Score 479; DB 15; Length 343;  
Best Local Similarity 37.9%; Pred. No. 2.4e-36;  
Matches 125; Conservative 54; Mismatches 121; Indels 30; Gaps 14;  
QY 18 GILALEKCFGSKNEQKEEDSKMYKRYREIVSVSLPSNDYWGDTMLYKGFWMQGYLVFGI 77  
Db 18 GTLAAAE--ARSPLPKEE-----FGDLVAALPRKEQYLDG-RLYEGFWLPEHYAPGI 66  
QY 78 MAFEDNFKAREVDIILTLPKAGTTWKALTETAILTRDVNHPSPHPLFPNPHSCYQN 137  
Db 67 IAFRRRTPRADVVVLASYPKCGTTLKALAFANTR-AAYPAD-EHPLLRLNPHDVIPF 124  
QY 138 LEYLYN-GRENTMPDLDMLNESPRLFAGHIPYSLLPASVYK----SGTKIINISNRKST 192  
Db 125 VEDVFTDGEH--AKLDML-PSPLRLNTHTPYQLLPESVWAGDGGGCKVYVICRDPKM 180  
QY 193 FVSWFKGNLNPDKLLDLEKSDVDFASGISFGPEWNPQAEFTNAA-STNSNLLILSYE 251  
Db 181 VVSLYHFMERLQPD--LSLAGVYESVADGTVPFGPMWHDHILGYWRASVSRPDRVLLRYE 238  
QY 252 EMLEKEV--ENVYKLAEFMGCGFTDDEEKQGIWDEIVKLCSPDNLKNQVKNKGS--SYN 307  
Db 239 DLLRDGAAGEHVRAVAFVGRFSAEEAAGAVASVVELCSFERMKALEVNERGTAGSYK 298  
QY 308 SKIDNKHFFPKRGKGEVDMANYLTSEMIX 337  
Db 299 S-MPRDAFFKGVAGDVAHNSMSPETAARLD 327

Search completed: March 20, 2004, 01:40:31

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 20, 2004, 03:02:04 ; Search time 5120 Seconds  
(without alignments)  
3309.996 Million cell updates/sec

Title: US-09-854-122-16

Perfect score: 2059

Sequence: 1 TRGITGIAVACLIMAGIL.....CLLSYSFVINNFILKKKK 391

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DRV=xlh  
-Q=/cgn2.1/USPTO.spool/US09854122/runat.18032004.100000.3253/app.query.fasta.1.583  
-DB=GenEmbl -CPWT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pcr -THE\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09854122@cgn.1.1.3508@runat.18032004.100000.3253 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DPSLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*

29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pin.\*  
35: em.htg.rod.\*  
36: em.htg.mam.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1996	96.9	1192	6	AX338910 Sequence
2	647.5	31.4	111989	2	AC135566 Medicago
C 3	620.5	30.1	88010	8	AL138649 Arabidops
C 4	620.5	30.1	89214	8	AL35992 Arabidops
C 5	594	28.8	138586	8	AC144726 Medicago
C 6	593	28.8	166490	2	AC146862 Medicago
C 7	587	28.5	107440	2	AC119410 Medicago
C 8	579	28.1	138586	8	AC144726 Medicago
C 9	578.5	28.1	102183	8	AC027134 Arabidops
C 10	578.5	28.1	103157	8	AC011810 Arabidops
C 11	574	27.9	108387	8	AC016662 Arabidops
C 12	573	27.8	1347	8	AY087493 Arabidops
C 13	572.5	27.8	1212	8	AY085411 Arabidops
C 14	570	27.7	1221	8	AY042887 Arabidops
C 15	567.5	27.6	113531	2	AC146342 Medicago
C 16	567	27.5	1041	8	AX697959 Sequence
C 17	567	27.5	1072	8	BT005398 Arabidops
C 18	567	27.5	1151	8	AX117408 Arabidops
C 19	567	27.5	1247	8	AX117463 Arabidops
C 20	567	27.5	44879	8	AC036238 Arabidops
C 21	566	27.5	1281	8	AY088081 Arabidops
C 22	566	27.5	108387	8	AC016662 Arabidops
C 23	564.5	27.4	1315	8	AX058698 Oryza sat
C 24	563.5	27.4	1092	8	AY081540 Arabidops
C 25	562.5	27.3	1017	6	AX697955 Sequence
C 26	559	27.1	1084	8	BT004984 Arabidops
C 27	558.5	27.1	1053	6	AX697957 Sequence
C 28	555.5	27.0	1077	6	AX080458 Sequence
C 29	555.5	27.0	1080	8	BT008847 Arabidops
C 30	555.5	27.0	1273	8	AY084999 Arabidops
C 31	555.5	27.0	1347	8	AY09809 Arabidops
C 32	555.5	27.0	86380	8	AB010697 Arabidops
C 33	553.5	26.9	121720	8	AC021044 Arabidops
C 34	549.5	26.6	1270	8	AY054219 Arabidops
C 35	548.5	26.6	1090	8	BT002108 Arabidops
C 36	548.5	26.6	1154	8	AY087921 Arabidops
C 37	548.5	26.6	1215	8	AY092961 Arabidops
C 38	548.5	26.6	87543	8	AC006836 Arabidops
C 39	546	26.5	1994	8	AF000307 Brassica
C 40	545	26.5	996	6	AX506332 Sequence
C 41	545	26.5	996	8	AF325065 Arabidops
C 42	542.5	26.3	1147	8	AY442306 Brassica
C 43	542.5	26.3	90341	8	AC006232 Arabidops
C 44	542.5	26.3	118196	8	AC005824 Arabidops
C 45	542	26.3	1179	8	AF375458 Arabidops

# ALIGNMENTS

RESULT 1

AX338910  
LOCUS AX338910 1192 bp DNA linear PAT 09-JAN-2002  
DEFINITION Sequence 15 from Patent WO0185971.  
ACCESSION AX338910  
VERSION AX338910.1 GI:18129074  
KEYWORDS  
SOURCE Zostera marina  
ORGANISM Zostera marina  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Zosteraceae; Zostera.  
REFERENCE 1  
AUTHORS Alberte, R.S. and Smith, R.D.  
TITLE Transgenic plants incorporating traits of Zostera marina  
JOURNAL Patent: WO 0185971-A 15 NOV-2001;  
Phycogen, Inc. (US)  
FEATURES  
source Location/Qualifiers  
1..1192  
/organism="Zostera marina"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:29655"  
join(1..33,37..1041,1045..1059,1063..1077,1081..1086,  
1090..1119,1123..>1191)  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAD20565.1"  
/db\_xref="GI:18129075"  
/translation="TRGITGAVACLELIMAGILALEKCFKSKNEQKEBDSKWKRY  
REIVSLPSNDYWGDTWELYGFWQMGVLPVGINAFEDNFKARETDIILTLPKAGTT  
WTKALTPALLRVNHPSPHLLFFNPGHCVQNLVLYMGRENTWPDLDMLNESPR  
LPAGHIPYSLIPASVLSKGTIKINISRNKSTFVSFWKFGNLINPDKLDELKGVDFI  
ASGISFCPEWFOHEFTNASTNSNLLLSYEEMLEKPVENVKKLAEFWGCGFTDDE  
EKQGVIVDLVKLCPDNKQOVNKNSSYNKIDNKHFFPKGVRVDRWANYLISEMIK  
KLETAGKINESEKHLNKNLTNLFNLNNYCEKSNISLCILSYSEVINNFILLKKKKK"

Alignment Scores:  
Pred. No.: 4,448-161 Length: 1192  
Score: 1996.00 Matches: 391  
Percent Similarity: 98.49% Conservative: 0  
Best Local Similarity: 98.49% Mismatches: 6  
Query Match: 96.94% Indels: 6  
DB: 6 Gaps: 6

US-09-854-122-16 (1-391) X AX338910 (1-1192)

QY 1 ThrArgGlyIleThrGlyIleAlaValAlaCys---LeuProLeuIleMetAlaGlyIle 19  
DB 1 ACGCGGGGAATAACTGGAATCGCTGTTGCTTCTAGCTACCCTGATATGCTGGAATT 60  
QY 20 LeuAlaLeuGluLysCysPheGlySerLysAsnGluGlnGluLysGluGluAspSerLys 39  
DB 61 TTAGCTTTGGAGAAATGTTTCGGATCCAGAAATGAGCAAGAGAGAGAGATTCCAAA 120  
QY 40 MetTyrLysArgTyrArgGluIleValSerSerLeuProSerAsnAspTyrTrpGlyAsp 59  
DB 121 ATGTACAGAGATATAGAGAGATTGTTTCTTCTCCTCCCTCGAATGATTATGGGGGAT 180  
QY 60 ThrMetArgLeuTyrLysGlyPheTrpGlnMetGlyTyrLeuValProGlyIleMetAla 79  
DB 181 ACCATGAGGTGTACAGGAGATTTTGGCAATGGGATATCTTGTACCTGGTATCATGCT 240  
QY 80 PheGluAspAsnPhelYsAlaArgGluThrAspIleLeuThrThrLeuProLysAla 99  
DB 241 TTGCAAGATAATTTCAAGGCTCGAGAGCGACATATCTTACGACTCTTCCAAAGGCT 300  
QY 100 GlyThrThrTrpThrLysAlaLeuThrPheAlaIleLeuThrArgAspValAsnHisPro 119  
DB 301 GGAACGACATGAGAGGACGCTGAGCTTTGGCATCTTACACGAGATTTTACACCCCA 360  
QY 120 SerSerProThrHisProLeuPhePheAsnProHisSerCysValGlnAsnLeuGlu 339  
DB 361 TCATCACCGACACATCCACTTTTGTCTTCAACCCCTCAATTCGGTGTTCAAAATTGGAG 420

QY 140 TyrLeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsnGluSerPro 159  
DB 421 TATTTGTACATGGGTAGAGAAAATACGATGCGACACCTCGATATGTTGAATGAATCCCG 480  
QY 160 ArgLeuPheAlaGlyHisIleProTyrSerLeuLeuProAlaSerValLeuLysSerGly 179  
DB 481 AGTTGTTTGGCGGACATCCCATACCTTTTGTTCGGGCTCTGTTTGAATCGGA 540  
QY 180 ThrLysIleIleAsnIleSerArgAsnArgLysSerThrPheValSerPheTrpLysPhe 199  
DB 541 ACAAAATCATCAATATAAGCGCAACCGTAAGAGTACATTTGTCTCTTTTGGAAATTT 600  
QY 200 GlyAsnLeuIleAsnProAspLysLeuLeuAspLeuGluLysSerValAspIlePheAla 219  
DB 601 GCGAATCTGATTAAACCCGCAAGTTATTGGACCTTCGAAAGAGCGTTGATATCTTCGA 660  
QY 220 SerGlyIleSerPheCysGlyProGluTrpAsnPheGlnAlaGluPheThrAsnAlaAla 239  
DB 661 TCGGGAATCTCTTTTGGACCGCAATGGAATTTCCAAAGCGGAGTTCCCAATCGCGC 720  
QY 240 SerThrAsnSerAsnLeuLeuLeuSerTyrGluGluMetLeuGluLysProValGlu 259  
DB 721 TCTACTAATTCAACTTCTGTTGTTGAGTTACGAGAAATGTTTAGAGAGCCAGTTGAA 780  
QY 260 AsnValLysLysLeuAlaGluPheMetGlyCysGlyPheThrAspAspGluLysGln 279  
DB 781 AATGTGAAGAACTAGCTGAGTTTCATGGGATGCGGTTCCACAGCATGAGGAGAAACAA 840  
QY 280 GlyIleValAspGluIleValLysLeuCysSerPheAspAsnLeuLysAsnGlnVal 299  
DB 841 GGGATTTGTTGATGAGATAGTTAAACTTTGATCTTCGCAATCTGAAGAAATCAACAGGTG 900  
QY 300 AsnLysAsnGlySerSerTyrAsnSerLysIleAspAsnLysHisPhePheArgLysGly 319  
DB 901 AACAAAACGGATCAAGCTACATTCGAAATCGACACCAAGCAATTTCTTCAGGAAGGT 960  
QY 320 GluValArgAspTrpAlaAsnTyrLeuThrSerGluMetIleLysLysLeuThrAla 339  
DB 961 GAGCTGAGAGATTGGCAAACTATCTAACGTCGGAATGATTAAGAAATCGAGACGGCC 1020  
QY 340 GlyLysIleAsnGluSerGlu---LysHisLeuLeuSer---AsnLysAsnLeuThr--- 356  
DB 1021 GGAATAATTAATGATCAAGATTAAGCAATTTATTTATCTCGAATAAGAAATCTTACATGA 1080  
QY 357 AsnPhe---AsnLeuAsnAsnTyrCysGluLysSerAsn---IleSerLeuCysLeuLeu 374  
DB 1081 AACTTCTGAAATCTTAATAATTACTGTGAGAAATCGAACTAAATATCTCTTTGTTTATTA 1140  
QY 375 SerTyrSerPheValIleAsnAsnPheIleLeuLeuLysLysLysLysLys 391  
DB 1141 TCGTATTTCATTCGTAATAATAATTTTCATTTTGTAAAAAATAAAAAA 1191

RESULT 2  
AC135566 111989 bp DNA linear HTG 12-DEC-2003  
LOCUS Medicago truncatula clone mth2-22g8, WORKING DRAFT SEQUENCE, 3  
DEFINITION ordered pieces.  
ACCESSION AC135566  
VERSION AC135566.13 GI:39752715  
KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT.  
SOURCE Medicago truncatula (barrel medic)  
ORGANISM Medicago truncatula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
Medicago.  
REFERENCE 1 (bases 1 to 111989)  
AUTHORS Shaull, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,  
Cook, D., Kim, D. and Roe, B.A.  
TITLE Medicago truncatula BAC Clone mth2-22g8  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 111989)

AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,  
Cook,B., Kim,D. and Roe,B.A.  
TITLE Direct Submission  
JOURNAL Submitted (18-OCT-2002) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
OK 73019, USA  
REFERENCE 3 (bases 1 to 111989)  
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,  
Cook,B., Kim,D. and Roe,B.A.  
TITLE Direct Submission  
JOURNAL Submitted (12-DEC-2003) Department Of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Farrington Oval, Room 208, Norman,  
OK 73019, USA  
COMMENT On Dec 12, 2003 this sequence version replaced gi:39573775.  
----- Genome Center  
Center: Department Of Chemistry And Biochemistry  
The University Of Oklahoma  
Center code:UOKNOR  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* been provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequences as soon as it is available and  
\* the accession number will be preserved.  
\* 1 11239: contig of 11239 bp in length  
\* 11240 11339: gap of unknown length  
\* 11340 20610: contig of 9271 bp in length  
\* 20611 20710: gap of unknown length  
\* 20711 111989: contig of 91279 bp in length.  
\* Location/Qualifiers  
      1..111989  
          /organism="Medicago truncatula"  
          /mol\_type="genomic DNA"  
          /db\_xref="taxon:3880"  
          /clone="mth2-2298"  
          /clone\_lib="Medicago truncatula BAC library H2"  
  
FEATURES  
source  
ORIGIN  
Alignment Scores:  
Pred. No.: 1,01e-43 Length: 111989  
Score: 647.50 Matches: 136  
Percent Similarity: 60.00% Conservative: 74  
Best Local Similarity: 38.86% Mismatches: 125  
Query Match: 31.45% Indels: 15  
DB: 2 Gaps: 9  
  
US-09-854-122-16 (1-391) x AC135566 (1-111989)  
QY 31 GluGlnGluLysGluGluAspSerLysMetTyrLysArgTyArgGluLeuValSerSer 50  
Db 43180 GATGAATCCATTGAAGAACAGCGTCAACTTAGCAGAAGAAACAAGAACTAATTTCTTC 43239  
QY 51 LeuProSerAsnAspTyrTrpGlyAspThr--MetArgLeuTyrLysGlyPheTrpGln 69  
Db 43240 CTTCCAAGAGAGAAAGGTGGAGACACCTTATATTCTATTTCCAAAGGATTTGGTGC 43299  
QY 70 MetGlyTyrLeuValProGlyIleMetAlaPheGluAspAsnPheLysAlaArgGluThr 89  
Db 43300 CAACACGCTGAATCCAAAGGCCATAACCCTTTCCAAAAGCATTTCCAAAGCTTAAGAAGT 43359  
QY 90 AspIleLeuThrThrLeuProLysAlaGlyThrThrTrpThrLysAlaLeuThrPhe 109  
Db 43360 GATGTTTTTGTCACCGTACCAAAATACAGGCACAACTTGTTAAAAAGCTCTTACCTAT 43419  
QY 110 AlaLeuThrArgAspValAsnHisProSerSerProThrHisProLeuLeuPhePhe 129  
Db 43420 GCCATTATGAATCGCCAAACCACTTTCATTCATCCAAAACCAATCCCTTTGCTAGTTTC 43479  
QY 130 AsnProHisSerCysValGlnAsnLeuGluTyrLeuTyrMetGlyValArgGluAsnThrMet 149

Coordinator: Marcel Salanoubat and Francis Quetier, Grouperment d'Interet Public, Centre National de Sequenage - GENOSCOPE; 2 rue Gaston Cremieux, BP191, 91006 Evry Cedex, France; <http://www.genoscope.cns.fr>

Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thai/>.

FEATURES

source 1..88010

/organism="Arabidopsis thaliana"

/mol\_type="genomic DNA"

/variety="Columbia"

/db\_xref="taxon:3702"

/chromosome="3"

1145..2116

/gene="T14D3.10"

/complement(1145..2116)

/gene="T14D3.10"

/complement(1145..2116)

/gene="T14D3.10"

/note="strong similarity to FLAVONOL 4'-SULFOTRANSFERASE - Flaveria chloraefolia, EMBL:M84136"

/codon\_start=1

/product="sulfotransferase-like protein"

/protein\_id="CAB72145.1"

/db\_xref="GI:6911845"

/db\_xref="GOA:Q9M1V2"

/db\_xref="SPTREMBL:Q9M1V2"

/translation="MENNRIEDLNERTKTLISLPSDKDFTGKTKICKYQGWTYHNV LQAVNPKFKQDPTDIIVASPKCGTTLKALTALLHRSKQPSHDDHPLSNRP HVLVPYFIDLYLSENPDLTKSSSPRLSTHVPSTLQEGKSGTKIVTSRVK DTLVYHFFKQTDKTLISSEFIDFEMFCRGVSTFGPTWDHVLVYWGSLDPNHV LFMFKEMKPEPDIQFAFLGCPFTKEEESGVDEIIDLCSLNNLSLEIKTGL KLSGRNKMFPKRGVGMKNTLTPMENKIDMIIOEKLQNSGLKF"

/complement(1145..2116)

/gene="T14D3.10"

/number=1

3483..4472

/gene="T14D3.20"

/complement(3483..4472)

/gene="T14D3.20"

/complement(3483..4472)

/gene="T14D3.20"

/complement(3483..4472)

/note="strong similarity to FLAVONOL 4'-SULFOTRANSFERASE - Flaveria chloraefolia, EMBL:M84136"

/codon\_start=1

/product="sulfotransferase-like protein"

/protein\_id="CAB72146.1"

/db\_xref="GI:6911846"

/db\_xref="GOA:Q9M1V1"

/db\_xref="SPTREMBL:Q9M1V1"

/translation="MDEKKITMNVNDELSESKTLISLPSDKNSGVNVCKYQGGW YTFPIQGVLPQKFKQDPTDIIVASPKCGTTLKALTALLHRSKQPSHDDHPL LSNPHVLSPLMXYLCSNPDLTKSSSRFLSTHVPSTLQEGKSGTKIVM SRNVDYLVSYHFFKQTDNIISSVDFEMFCRGVSTFGPTWDHVLVYWGSL EDPNHVLMFKEMKPEPDIQFAFLGCLFTKEEESGLVDEIIDLCSLNNLSLE INKTKLHSTGRNKMFPKRGVGMKNTLTPMENKIDMIIOEKLQNSGLKF"

/complement(3483..4472)

/gene="T14D3.20"

/number=1

5519..9216

/gene="T14D3.30"

/complement(5519..9216)

7034..7426,7615..7696,8280..8578,9067..9216))

/gene="T14D3.30"

/complement(5519..9216)

7034..7426,7615..7696,8280..8578,9067..9216))

/gene="T14D3.30"

/note="strong similarity to 2-phosphoglycerate kinase - Methanococcus Jannaschii, PIR:A64485

Contains ATP/GTP-binding site motif A (P-loop)

A164-171:Prokaryotic membrane lipoprotein lipid

attachment site A158-168:Prokaryotic membrane lipoprotein

lipid attachment site AA386-396"

/codon\_start=1

/product="putative protein"

/protein\_id="CAB72147.1"

/db\_xref="GI:6911847"

/db\_xref="SPTREMBL:Q9M1V0"

/translation="MWEATKVLIVVREEGDDDDNDDSDFRVTRPVLOSTLOLMGCK ARHAFKLSRRVVELIRSEGCSTPENGKEPEFAKVGSGSTCVKLNCLVAGDVKN KSPFENYKRTTIVSREIFVDVCCDAEYIKYVGRDQADLILACIRIERKESVTV LDCGTCGKSTISALIGSLGITVTVSDTISRMRSPFADEKONPILWASTTHAGY LDPVAEASAKAKKAKKIGSRGVNAGNAQKTDSNSSTTELLSHKQMAIEGYKQSE MVIDSLDLTLTWEERNESVVVEGHLNFMVGLMKKHPISIVPFMYIANEKHLER FAYPAKYMTLDPAKVKYKIRNRTIQLYCKRADKHLVKNNTNIVDSKVATIHAT VFGCLRRTEGKLYDTTNTVSVIIDEHNOCAANSLTSKGMFOVIOGSGRRFWA LCNTDGTAKTVASVGKIKPVNTEMDDGTEHQLKAEVNLQFHFHGISAWPDS GATSHAGSVDLRLADIETSRHRYSCSSPRISDDEYDKLMEEQSVNGSDEDEEDG DDFHEPDSDELDNDRNDRDEIGSVDEESTKDEYDDEEDKEDKQSVNGSDEDEEDG DTISWVONNHEASTKNDKYSQNLDFLKTNTQPLTESLELTSTSYRNRMGVAASD KAKMRKSLSPVPVKGHSIIDQILANQTDVSL"

/complement(5519..9216)

/gene="T14D3.30"

/number=1

/complement(5903..6006)

/gene="T14D3.30"

/number=1

/complement(6007..6129)

/gene="T14D3.30"

/number=2

/complement(6130..6216)

/gene="T14D3.30"

/number=2

/complement(6217..6936)

/gene="T14D3.30"

/number=3

/complement(6937..7033)

/gene="T14D3.30"

/number=3

/complement(7034..7426)

/gene="T14D3.30"

/number=4

/complement(7427..7614)

/gene="T14D3.30"

/number=4

/complement(7615..7696)

/gene="T14D3.30"

/number=5

/complement(7697..8279)

/gene="T14D3.30"

/number=5

/complement(8280..8578)

/gene="T14D3.30"

/number=6

/complement(8579..9066)

/gene="T14D3.30"

/number=6

/complement(9067..9216)

/gene="T14D3.30"

/number=7

14402..14473

/note="72 bp long terminal repeat"

14495..14861

/note="similarity to fragments of copia-like polyprotein"

19002..19073

/note="72 bp long terminal repeat"

19415..21625

/gene="T14D3.40"

Join(19415..19534,19838..20410,20508..20662,20758..20837, 20930..21066,21169..21272,21360..21450,21533..21625)

/gene="T14D3.40"

/note="strong similarity to n-acetylglucosaminyl-phosphatidylinositol - Schizosaccharomyces pombe, PIR:740367"

/codon\_start=1







```

/number=4
25629..25721
/genes="F14D17_40"
/number=5
25722..25836
/genes="F14D17_40"
/number=5
25837..25843
/genes="F14D17_40"
/number=6
25844..25967
/genes="F14D17_40"
/number=6
25968..26074
/genes="F14D17_40"
/number=7
26075..26338
/genes="F14D17_40"
/number=7
26339..26563
/genes="F14D17_40"
/number=8
32738..33813
/genes="F14D17_50"
join(32738..32746,32854..32996,33276..33354,33471..33624,
33686..33813)
/genes="F14D17_50"
/codon start=1
/product="putative protein"
/protein_id="CAB89413.1"
/db_xref="GI:7671399"
/db_xref="SPTRMBL:O9LX7"
/translation="MAELSSGKSTLASIANKMYOFARKIBETDFLLSPNSYKDSI
DYTSLELRRLRWQDTRAIADICNIRVKGDSQNGKIEPKWSRKIHPLLSNR
KYWAENELRRALLQRGDSIAEKSISIGKDSRYEREVPTPASFSGTGVGEIRDSFF

Alignment Scores:
Pred. No.: 1,56e-41 Length: 89214
Score: 620.50 Matches: 143
Percent Similarity: 57.55% Conservative: 78
Best Local Similarity: 37.24% Mismatches: 126
Query Match: 30.14% Indels: 37
DB: 8 Gaps: 11

US-09-854-122-16 (1-391) x ATF14D17 (1-89214)

Qy 25 CysPheGlySerLysAenGluGlnGluLysGluGlu-AspSerLysMetTyrLysArgTy 44
Db 81424 TGTAAATGGATGAGAAAGAGATTACGATGAACGTGAGAAACGATGAGTTAAGCGAAGATC 81365
Qy 44 rArgGluIleValSerSerLeuProSerAsn--AspTyr-TripGlyAspThrMetArgLe 63
Db 81364 AAAAATCTAATCTCTTCACTCTCTTCAGCAAAAATTCACCTGGGTAAATGTTGTAA 81305
Qy 63 uTyLysGlyPheTrpGlnMetGlyTyrLeuValProGlyIleMetAlaPheGluAspAs 83
Db 81304 GTACCAAGGATGTGGTATACCTCGCGCTATTTCTCAAGGTGTCTCAATTTCCAGAAAA 81245
Qy 83 nPheIleValArgGluThrAspIleLeuThrThrLeuProLysAlaGlyThrThrTr 103
Db 81244 CTTTAAGCCTCAAGACACCGATATCATGTGTCTTCTTCAATTCGCGCACCACTTG 81185
Qy 103 pThrLysAlaLeuThrPheAlaIleLeuThrArgAspValAsnHisProSer---SerPr 122
Db 81184 GCTCAAGCGCTTACATTCGCACCTGTCGTAGATCG---AAGCACCTTCTCATGATGA 81128
Qy 122 oThrHisProLeuLeuPhePheAenProHisSerCysValGlnAsnLeuGlu---TyrLe 141
Db 81127 TCATCATCCACTCTTCTGTATATCCACAGTCTTTTACCCTCCCTTGGATGTATCT 81068
Qy 141 uTyMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsnGluSerProArgLe 161
Db 81067 CTAICTATGTAGCGAAAT-----CCCGACCTTACCAAGTCTCATCATCTTCGAGGCT 81014

```

```

161 uPheAlaGlyHisIleProTyrSerLeuLeuProAlaSerValLeuLysSerGlyThrLy 181
Db 81013 GTTTTCGACACACATGCGGTGCACATATGCAAGAGGTCCTCAAGGTTCTACTGTAA 80954
Qy 181 sIleAlaAsnIleSerArgAsnArgLysSerThrPheValSerPheTrpLysPhe----- 199
Db 80953 AATTGTGTATATGTCTAGAAACGTAAGACACATTTGGTTTCATATTTGGCATTTCTTTTG 80894
Qy 200 -----GlyAsnLeuIleAsnProAspLysLeuLeuAspLeuGluLys 214
Db 80893 TAAGAAACAACCGATGATACATAATAGC-----AGTGTGAGGATAC 80849
Qy 214 rValAspIlePheAlaSerGlyIleSerPheCysGlyProGluTrpAsnPheGlnAlaGl 234
Db 80848 ATTTCAGATGTTTGTAGGGAGTCAACTTTTTCGGCCCTTTTGGGACCATGTCCTTAAG 80789
Qy 234 uPheThrAsnAlaAlaSerThrAsnSerAsn---LeuLeuLeuLeuSerTyrGluGluMe 253
Db 80788 CTACTGGAGAGGAAGCTTGGAGATCCAAACCATGCTCTTTATGAGTTTGGAGGAT 80729
Qy 253 tLeuGluLysProValGluAsnValLysLysLeuAlaGluPheMetGlyCysGlyPheTh 273
Db 80728 GAAAGAGAACCTCGTCAGCAGATCAAGAGACTAGCGAATTCCTAGGATGCTCTTTTAC 80669
Qy 273 rAspAspGluGluLysGlnGlyIleValAspGluIleValLysLeuCysSerPheAspAs 293
Db 80668 TAAGAGAGAGAGAAACGCGCTTGGTGATGAGATTATCGATCTTTGCTCTCTACGTAA 80609
Qy 293 nLeuLysAsnGlnGlnValAsnLysAsnGlySerSerTyrAsnSerLysLysLeuAspAsnLy 313
Db 80608 TCTGAGCAGTTTGGAGATCAATAAGACCGGAAATTCATAGTAGTACTGGTAGAGAGACAA 80549
Qy 313 sHisPhePheArgLysGlyGluValArgAspTTPAlaAsnTyrIleThrSerGluMetIl 333
Db 80548 AACATTTTTCGTAAGGAGAGTTGGTACTGGAAGAACTATTTGACTCTCTGAAATGGA 80489
Qy 333 eLysLysLeuGluThrAlaGlyLysIleAsnGluSerGluLysHisLeuLeuLysL 353
Db 80488 GAACAAAATCGAC-----ATGATCATTTCAAGA 80462
Qy 353 sAsnLeuThrAsnPheAsnLeuAsnAsn---TyrCysGluLysSerAsnIleSerLeu-C 372
Db 80461 GAAACTTCAAAACTCTGCTGTTGAAATTCGAGTTGTGTGCTTTGAGCTTTTCTTAAT 80402
Qy 372 yLeu-----LeuSerTyrSerPheValIleAsnAsnPheIleLeuLeuLysL 388
Db 80401 GCAITGTAATAATAGACTACTTTTATGTTGTTATATTCCTTCTGTATGCTCTCGT 80342
Qy 388 yLeuLys 390
Db 80341 GTAAGCGT 80334

RESULT 5
AC144726/c
LOCUS AC144726 138586 bp DNA linear PLN 30-AUG-2003
DEFINITION Medicago truncatula clone mth2-7k13, complete sequence.
ACCESSION AC144726
VERSION AC144726.6 GI:34365847
KEYWORDS HTG.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE 1 (bases 1 to 138586)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
TITLE Medicago truncatula BAC Clone mth2-7k13
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 138586)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,

```

```

Cook, D., Kim, D. and Roe, B.A.
Direct Submission
Submitted (14-MAY-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS
3 (bases 1 to 138586)
Shauli, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,
Cook, D., Kim, D. and Roe, B.A.
Direct Submission
Submitted (15-JUL-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS
4 (bases 1 to 138586)
Shauli, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,
Cook, D., Kim, D. and Roe, B.A.
Direct Submission
Submitted (30-AUG-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT
On Aug 30, 2003 this sequence version replaced gi:32567849.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code: UOKNOR
-----
FEATURES
source
Location/Qualifiers
1..138586
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/db_xref="taxon:3880"
/clone="mth2-7k13"
/clone_lib="Medicago truncatula BAC library H2"
ORIGIN
Alignment Scores:
Pred. No.: 4,74e-39 Length: 138586
Score: 594.00 Matches: 138
Percent Similarity: 58.79% Conservative: 76
Best Local Similarity: 37.91% Mismatches: 133
Query Match: 28.85% Indels: 17
DB: 8 Gaps: 9
US-09-854-122-16 (1-391) x AC144726 (1-138586)
QY 34 LysGluLysSerLysMetTyrLysArgTyrArgGluLeuValSerSerLeuProSer 53
Db 33772 AAAGAAGATGAC-----TTAGGGCAAGATGCAAGGACTTGATACAAACCTTGCCATA 33719
QY 54 AsnAspTyrTrpGlyAsp---ThrMetArgLeuTyrLysGlyPheTrpGlnMetGlyTyr 72
Db 33718 GTAGAGGGTGGATCGATCCACCTTTTCATGAATATCAAGGATTTGGTTGCTCCCAAG 33659
QY 73 LeuValProGlyLeuMetAlaPheGluAspAsnPheLysAlaArgGluThrAspIleLeu 92
Db 33658 ATTCTCCAGGTGCTTTATCTTTTCAAAAGACCTTTCAGGCCATTGATACAGATATATC 33599
QY 93 LeuThrThrLeuProLysAlaGlyThrTrpThrLysAlaLeuThrPheAlaLeuLeu 112
Db 33598 TTAGTACAAATCCCAATCAGGTACCATCTGCTCAAGCATTTGATTTGCTTTGATA 33539
QY 113 ThrArgAspValAsnHisProSerProThrHisProLeuLeuPhePheAsnProHis 132
Db 33538 AACCGTAACAATATCCAAATATTATCAAAAAATCATCTTTTACTCATCTCCACCCCTCAT 33479
QY 133 SerCysValGlnAsnLeuGluTyrLeuTyrMetGlyArgGluAsnThrMetProAspLeu 152
Db 33478 GTTCTGTTCTTTCATGGAGATTAACTTTACTATGAACTGATATTGTTCTGCACTC 33419
QY 153 AspMetLeuAsnGluSerProArgLeuPheAlaGlyHisIleProTyrSerLeuLeuPro 172
Db 33418 AATTCACATATCC---CTTCCAAAGAAATCTTCTCCACACACATCTTATGTTTGTGGCA 33362
QY 173 AlaSerValLeuLysSerGlyThrLysIleLeuAsnIleSerArgAsnArgLysSerThr 192

```

```

Db 33361 AAATCTTTGAAAGAGTCCAACTGTAAGTGCTGTATCTTTGTAGAGATCCTTAAGACACT 33302
QY 193 PheValSerPheTrpLysPheGlyAsnLeuIleAsnProAspLys-----LeuLeuAsp 210
Db 33301 TTGTGTTTCACTTGGCATTTCTCAACAAGCTAAGAGACAAAGTAGAGGACACTTCCA 33242
QY 211 LeuGluLysSerValAspIlePheAlaSerGlyIleSerPheCysGlyProGluTrpAsn 230
Db 33241 TTAGAAGAGCGCTTTTGAGAGTTTGTAGAGAACACTTCTTATGAGACCGTTTGGGAT 33182
QY 231 PheGlnAlaGluPhe---ThrAsnAlaAlaSerThrAsnSerAsnLeuLeuLeuSer 249
Db 33181 CATGTATTAGTACTCTGGAAGAAAGCTTGAAGACCGGAGAGGTTATGTTTGTAA 33122
QY 250 TyrGluGluMetLeuGluLysProValGluAsnValLysLysLeuAlaGluPheMetGly 269
Db 33121 TATGAAGAAATGAAATGAAACCTAATTTGTTTGAAGAGAGATTGCTAAGTTTGGGG 33062
QY 270 CysGlyPheThrAspAspCysGlyGlnGlyIleValAspGluIleValLysLeuCys 289
Db 33061 TGTCCATTCTCGAAGAAAGAAATCCATGTTGTTGTTGATGATATATTACATTTGCT 33002
QY 290 SerPheAspAsnLeuLysAsnGlnValAsnLysAsnGlySerSerTyrAsnSerLys 309
Db 33001 AGTTTGTAGAGCTTAGCAACTTGGAGATCAATAAGAACGGG---AAAAATTTGTCGGAA 32945
QY 310 IleAspAsnLysHisPhePheArgLysGlyGluValArgAspTrpAlaAsnTyrLeuThr 329
Db 32944 ATTGAACAACAAGAACTTTTCCGCTCTGGCAAGTTGGAGATTGGAANAATTTCTCACA 32885
QY 330 SerGluMetIleLysLysLeuGluThrAlaGlyLysIleAsnGluSerGluLysHisLeu 349
Db 32884 ACTAAATGTTGAGCAGTTAAACATAGTTGCGAA---GAGAAATTTGCTAAACATGGG 32828
QY 350 LeuSer-----AsnLysAsnLeuThrAsnPheAsn---LeuAsnAsnTyr----- 363
Db 32827 TTGAGTTTGTAGATTTTGAAATCTCTGTTCACTTTATGGCTGATGGCTTTCTTAAT 32768
QY 364 ---CysGluLysSerAsnIleSerLeuCysLeuLeuSerTyrSerPheValIleAsnAs 382
Db 32767 CCTTGTACCTTTCTTTTTCAGCTTTGTTCTGTTGGGAATCATTTGTTCTTCTGGAT 32708
QY 382 nPheIleLeu 385
Db 32707 ATTCTTTTG 32698
RESULT 6
AC146862 166490 bp DNA linear HTG 05-DEC-2003
LOCUS Medicago truncatula clone mth2-174p12, WORKING DRAFT SEQUENCE, 26
DEFINITION unordered pieces.
ACCESSION AC146862
VERSION AC146862.8 GI:38708066
KEYWORDS HTG: HTGS PHASE1; HTGS DRAFT.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE
1 (bases 1 to 166490)
AUTHORS Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B., Cook, D., Kim, D.
and Roe, B.A.
TITLE Medicago truncatula BAC Clone mth2-174p12
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 166490)
AUTHORS Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B., Cook, D., Kim, D.
and Roe, B.A.
TITLE Direct Submission
Submitted (15-OCT-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA

```

REFERENCE 3 (bases 1 to 166490)  
AUTHORS Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.  
and Roe,B.A.  
TITLE Direct Submission  
JOURNAL Submitted (05-DEC-2003) Department of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Farrington Oval, Room 208, Norman,  
OK 73019, USA

## COMMENT

On Dec 5, 2003 this sequence version replaced gi:38678574.

----- Genome Center  
Center: Department Of Chemistry And Biochemistry  
The University Of Oklahoma  
Center code:UOKNOR

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 26 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 2098: contig of 2098 bp in length  
\* 2099 2198: gap of unknown length  
\* 2199 5183: contig of 2985 bp in length  
\* 5184 5283: gap of unknown length  
\* 5284 7288: contig of 2005 bp in length  
\* 7289 7388: gap of unknown length  
\* 7389 9418: contig of 2030 bp in length  
\* 9419 9518: gap of unknown length  
\* 9519 11543: contig of 2025 bp in length  
\* 11544 11643: gap of unknown length  
\* 11644 14312: contig of 2669 bp in length  
\* 14313 14412: gap of unknown length  
\* 14413 17259: contig of 2847 bp in length  
\* 17260 17359: gap of unknown length  
\* 17360 19862: contig of 2503 bp in length  
\* 19863 19962: gap of unknown length  
\* 19963 22492: contig of 2530 bp in length  
\* 22493 22592: gap of unknown length  
\* 22593 25048: contig of 2456 bp in length  
\* 25049 25148: gap of unknown length  
\* 25149 27970: contig of 2822 bp in length  
\* 27971 28070: gap of unknown length  
\* 28071 30512: contig of 2442 bp in length  
\* 30513 30612: gap of unknown length  
\* 30613 33341: contig of 2729 bp in length  
\* 33342 33441: gap of unknown length  
\* 33442 35670: contig of 2229 bp in length  
\* 35671 35770: gap of unknown length  
\* 35771 40178: contig of 4408 bp in length  
\* 40179 40278: gap of unknown length  
\* 40279 42682: contig of 2404 bp in length  
\* 42683 42782: gap of unknown length  
\* 42783 45170: contig of 2388 bp in length  
\* 45171 45270: gap of unknown length  
\* 45271 48758: contig of 3488 bp in length  
\* 48759 48858: gap of unknown length  
\* 48859 52749: contig of 3891 bp in length  
\* 52750 52849: gap of unknown length  
\* 52850 62432: contig of 9583 bp in length  
\* 62433 62532: gap of unknown length  
\* 62533 70569: contig of 8037 bp in length  
\* 70570 70669: gap of unknown length  
\* 70670 86794: contig of 16125 bp in length  
\* 86795 86894: gap of unknown length  
\* 86895 100829: contig of 13935 bp in length  
\* 100830 100929: gap of unknown length  
\* 100930 115788: contig of 14859 bp in length  
\* 115789 115888: gap of unknown length  
\* 115889 143040: contig of 27152 bp in length  
\* 143041 143140: gap of unknown length  
\* 143141 166490: contig of 23350 bp in length.

Location/Qualifiers

source 1. 166490  
/organism="Medicago truncatula"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:3880"  
/clone="mth2-174p12"  
/clone\_lib="Medicago truncatula BAC library H2"

## ORIGIN

Alignment Scores: 7.14e-39 Length: 166490  
Pred. No.: 593.00 Matches: 133  
Score: 56.06% Conservative: 66  
Percent Similarity: 37.46% Mismatches: 142  
Best Local Similarity: 28.80% Indels: 14  
Query Match: 2 Gaps: 8  
DB: 2  
US-09-854-122-16 (1-391) x AC146862 (1-166490)  
Qy 9 ValAlaCysLeuProLeuIleMetAlaGlyIleLeuAlaLeuGluLysCysPheGlySer 28  
Db 55772 GTGTGGTGTCTCTTCATGGCTTCAACAGATCTCACAATATGAAGAGACCAATCA 55831  
Qy 29 LysAsnGluGlnGluLys---GluGluAspSerLysMetTyrLysArgTyrArgGluIle 47  
Db 55832 AGGATGACACAGAAACAAATTAGCGAAGAAACAGAGTAATCCATGATTATAACCTA 55891  
Qy 48 ValSerSerLeuPro---SerAsnAspTyrTyrGlyAspThrMetArgLeuTyrLysGly 66  
Db 55892 ATCTGTGCTCTCTCTAGAGAAATGGATGTGATCTCAATATTTTATTTTATCTGGA 55951  
Qy 67 PheTyrGlnMetGlyTyrLeuValProGlyIleMetAlaPheGluAspPheLysAla 86  
Db 55952 TTTTGTGTCCATCACTCTTATCCAACTCTTTCAAAATAATTTTCATGCC 56011  
Qy 87 ArgGluThrAspIleLeuThrLeuProLysAlaGlyThrThrThrThrLysAla 106  
Db 56012 AAAGATAGTATATTGTGTTGTCATCCATCCAAATCTGGCACTACTTGGCTTAAGGT 56071  
Qy 107 LeuThrPheAlaIleLeuThrArgAspValAsnHisProSerSerProThrHisProLeu 126  
Db 56072 CTTGCATATGCTATTGTTAATCGCAACATTTTACCTCTTAGAGAATAACCATCCATTA 56131  
Qy 127 LeuPhePheAsnProHisSerCysValGluAsnLeuGluTyrLeuTyrMetGlyArgGlu 146  
Db 56132 CTATCATTTAATCCACATGAACCTTGTCCTCATTTTGAAGTAAGTAAATATACAGAT 56191  
Qy 147 AsnThrMetProAspLeuAspMetLeuAsn---GluSerProArgLeuPheAlaGlyHis 165  
Db 56192 GGTCAAATGCTCAAATGTATGTACCAAATATGGTTCAACCAAGCTTTTATGGAACCTCAC 56251  
Qy 166 IleProTyrSerLeuLeuProAlaSerValLeuLysSerGlyThrLysIleIleAsnIle 185  
Db 56252 ATGCCATTTCTCTCGTGGCCCAAGTCAATTCAGAGTCCAATTTGTAATAATTTATATA 56311  
Qy 186 SerArgAsnArgLysSerThrPheValSerPheTyrLysPheGlyAsnLeuIleAsnPro 205  
Db 56312 TGTAGAACCCTATTGTATACATTTGTATCATATTTGGATTTTCATCAACAACCTCAGATTA 56371  
Qy 206 AspLysLeu-----LeuAspLeuGluLysSerValAspIlePheAlaSerGlyIle 222  
Db 56372 AAGAAATCTTTAAGTGAATTAACCTTTGGAGGAATCTTTTGAAGGTACTCTGAAGGGAATA 56431  
Qy 223 SerPheCysGlyProGluTyr---AsnPheGluAlaGluPheThrAsnAlaAlaSerThr 241  
Db 56432 TGTTTTTTGTCTCGTCTGGACAATATGTTGGGTACTTGAAGGAGACGATAGAAAGA 56491  
Qy 242 AsnSerAsnLeuLeuLeuSerTyrGluGluMetLeuGluLysProValGluAsnVal 261  
Db 56492 CCAGATAGGGTTTTTATCTTTGAAATATGAAGACCTTAAGAGAGATGTTAATTTTTCACACG 56551  
Qy 262 LysLysLeuAlaGluPheMetGlyCysGlyPheThrAspAspGluGluLysGlnGlyIle 281  
Db 56552 AAAAGAAATTGCAGAATTTGTGGGATTTCTCTTACTGAAGAGAGGAAATAATGGAAGT 56611



```

QY 261 ValLysLeuLeuAlaGluPheMetGlyCysGlyPheThrAspAspGluGlySerGlnGly 280
Db 22107 ACGAAGAAATTCAGAAATTTGGGATTCCTTCATCTGAGAGAGAGAAATAATGGA 22048
QY 281 IleValAspGluIleValLysLeuCysSerPheAspAsnLeuLysAsnGlnValAsn 300
Db 22047 GTGATTGAAATATAATCAAGTATATGACGCTTTGAGAGTATGAGGAATCAAGTGAAAT 21988
QY 301 LysAsnGlySerSerTyRAsnSerLysIleAspAsnLysHisPheArgLysGlyGlu 320
Db 21987 AAATCGGAACATGCA---GCATTAAGATTGAGAGAGATTTTCTTCGAAAGGAGAA 21931
QY 321 ValArgAspTrpAlaAsnTrpLysLeuThrSerGluMetIleLysLysLeuGluThrAlaGly 340
Db 21930 ATAGGGGATGGTGGAATATCTCTCCCTTCGATGATATAAAATTA-----TCC 21880
QY 341 LysIleAsnGluSerGluLysHisLeuSerLysAsnLeuThr 356
Db 21879 AAAGTCATGGAAGAAAA-----TTAAGTGATCAAGCCTATCA 21841

RESULT 8
AC144726 138586 bp DNA linear PLN 30-AUG-2003
LOCUS Medicago truncatula clone mth2-7k13, complete sequence.
DEFINITION AC144726
ACCESSION AC144726.6 GI:34365847
VERSION HTG.
KEYWORDS Medicago truncatula (barrel medic)
SOURCE Medicago truncatula
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots;
Rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE 1 (bases 1 to 138586)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
TITLE Medicago truncatula BAC Clone mth2-7k13
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 138586)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (14-MAY-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 3 (bases 1 to 138586)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (15-JUL-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 4 (bases 1 to 138586)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (30-AUG-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On Aug 30, 2003 this sequence version replaced gi:32567849.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----
FEATURES
source
1. 138586
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/db_xref="taxon:3880"
/clone="mth2-7k13"
/clone_lib="Medicago truncatula BAC library H2"

ORIGIN

```

```

Alignment Scores:
Pred. No.: 9,04e-38 Length: 138586
Score: 579.00 Matches: 138
Percent Similarity: 58.33% Conservative: 72
Best Local Similarity: 38.33% Mismatches: 131
Query Match: 28.12% Indels: 20
DB: Gaps: 9

US-09-854-122-16 (1-391) x AC144726 (1-138586)
QY 36 GluAspSerLysMetTyRAsnGlyArgGluValSerSerLeuProSerAsnAsp 55
Db 38384 CAAGATGATGACATTAAGCCAAAGATGCAAGAGATGATACCCCTTGCCATCGAAAAA 38443
QY 56 TyRTrpGlyAspThrMetArgLeu---TyRLeGlyPheTrpGlnMetGlyTyRLeuVal 74
Db 38444 GCCTGGATTTTAAACCCACACCTATCAATATCAAGGATTTGGATTGCCACAAAAAATTT 38503
QY 75 ProGlyIleMetAlaPheGluAspAsnPheLysAlaArgGluThrAspIleLeuThr 94
Db 38504 CAAGGTGTTTATCTGTCTCAGAAGCACCTTTCAAGCCCTAGATTTCAGATATCATCTAGTT 38563
QY 95 ThrLeuProLysAlaGlyThrThrTrpThrLysAlaLeuThrPheAlaIleLeuThrArg 114
Db 38564 ACCGTCCTCCAAATCAGGTACCACTTGGCTCAAGCATTGACTTTGCTTTGCTAAACCGC 38623
QY 115 AspValAsnHisProSerSerProThrHisProLeuLeuPhePheAsnProHisSerCys 134
Db 38624 AACAAATATCCAAATATTATCAATAACCACTCTTGTCTCAATTTCCAAATCCCAATCTCTT 38683
QY 135 ValGlnAsnLeuGluTyRLeuTyRMetGlyArgGluAsnThrMetProAspLeuAspMet 154
Db 38684 GTTCCCTTCTTGAGATTGACCTTTACAGTGATAAGACTATGTTCTCTCACTTAACCTACT 38743
QY 155 LeuAsnGluSerProArgLeuPheAlaGlyHisIleProTyRSerLeuLeuProAlaSer 174
Db 38744 TTATCTCAT---CCAGACTCTTCTCCACACATATCTCTTGTGCTTTGTTGCCAAATCT 38800
QY 175 ValLeuLysSerGlyThrLysIleAsnIleSerArgAsnArgLysSerThrPheVal 194
Db 38801 GTGAAGATCAACTTCAGAGTTGTATCTGTGTAGAGATCCTTAAGACATGTTTGT 38860
QY 195 SerPhe-----TrpLysPheGlyAsnLeuIleAsnProAspLysLeu-----LeuAsp 210
Db 38861 TCACT-TTACTTTGGCATTTTCAAAACAAGTAAAGACACAGATAGACCACTTCCA 38919
QY 211 LeuGluLysSerValAspIlePheAlaSerClyIleSerPheCysGlyProGluTrp--- 229
Db 38920 TTAGAGAGGCTTTTGAGAGTTTGTAGAGGAGTGAGTTCTTATGACCTTTTGGAG 38979
QY 230 AsnPheGlnAlaGluPheThrAsnAlaAlaSerThrAsnSerAsnLeuLeuLeuSer 249
Db 38980 CATGTATTAGGATATTGGAAGGAAACTTGGAAAGGCCAGAGAGGTGATGTTTATAAAA 39039
QY 250 TyrGluGluMetLeuGluLysProValGluAsnValLysLysLeuAlaGluPheMetGly 269
Db 39040 TTTGAAGAAATGAAATCAAACTAGTTTATTATTGAAAAGATGCTGAGTTTCTGGGT 39099
QY 270 CysGlyPheThrAspAspGluLysGlnGlyIleValAspGluIleValLysLeuCys 289
Db 39100 TGTCCATCTTTATGAAGAGATCCAGGCGCATGATCGATGATATATAAATTGTGT 39159
QY 290 SerPheAspAsnLeuLysAsnGlnValAsnLysAsnGlySerSerTyRAsnSerLys 309
Db 39160 AGTTTGAAGAACTTAAGCAACTTGGAGGTCAACAGACC---GTAAGAGTAGCAACAGGT 39216
QY 310 IleAspAsnLysHisPhePheArgLysGluValArgAspTrpAlaAsnTyRLeuThr 329
Db 39217 GTTGAAGAAATGAAGCTTTTTCGTCGTGGCCAAAGTTGGAGATTGGAAAAACCTTCTACA 39276
QY 330 SerGluMetIleLysLeuGluThrAlaGlyLysIleAsnGluSerGluLysHis--- 348

```











complete sequence.  
AC016662  
VERSION AC016662.7 GI:12325130  
KEYWORDS  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
REFERENCE 1 (bases 1 to 108387)  
AUTHORS Lin, X., Kaul, S., Town, C.D., Benito, M.-I., Creasy, T.H., Haas, B.J., Wad, D., Wai, R., Rong, C.M., Koo, H., Fujii, C.Y., Uterback, T.R., Barnstead, M.E., Bowman, C.L., White, O., Nierman, W.C. and Fraser, C.M.  
Arabidopsis thaliana chromosome 1 BAC F2P9 genomic sequence  
Unpublished  
2 (bases 1 to 108387)  
AUTHORS Lin, X. and Kaul, S.  
Direct Submission  
Submitted (04-DEC-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org  
3 (bases 1 to 108387)  
AUTHORS Town, C.D. and Kaul, S.  
Direct Submission  
Submitted (19-JAN-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org  
On Jan 19, 2001 this sequence version replaced gi:12280816.  
Address all correspondence to: at@tigr.org

BAC clone F2P9 is from Arabidopsis thaliana chromosome 1  
The orientation of the sequence is from SP6 to T7 end of the BAC clone.  
Genes were identified by a combination of several methods: Gene prediction programs including Genscan+ (Chris Burge, <http://CCR-081.mit.edu/GENSCAN.html>), GeneMarkEM (Mark Borodovsky, <http://genemark.biology.gatech.edu/GeneMark/>), GlimmerA (a variant of GlimmerM, see Mihaela Pertea,  
<http://www.tigr.org/softlab/glimmerm.htm>), and GeneSplicer (Mihaela Pertea and Steven Salzberg, contact [mpertea@tigr.org](mailto:mpertea@tigr.org)), searches of the complete sequence against a peptide database and the plant EST database at TIGR (<http://www.tigr.org/tdb/tgi.shtml>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).  
Location/Qualifiers  
1..108387  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/cultivar="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="1"  
/clones="F2P9"  
35..1255  
/gene="F2P9.1"  
join(<35..663,669..>1255)  
/gene="F2P9.3"  
join(<35..663,669..1255)  
/gene="F2P9.1"  
/codon\_start=2  
/product="unknown protein, 5' partial; 35-1255"  
/protein\_id="AAG52507.1"  
/db\_xref="GI:12325131"  
/translation="KRTLLANLLQRYGFPSSLOHFLSRNHLSSDLVETELISGL  
LSLIPKQSLVSLIDCPNVLRSFLKRWKRVPLSCNGHGVVSSAISKLSVHSSRIQ  
IGPKFNECVRLKSGFCDSVRLSSPFGVLYNLEIRKTEFLVGTGIARDNI  
ERFHFVPEVLGIGTETRLKPLLDDEFKMGFSGKDDVKYKIAERERFLVNSLPCLIE

INTLKREIVIRVSIISGAPAGPEVKLRVDCLCKGLIRRDFAKVVWKEPRVILVEI  
EDIEKKIEFLNRMGFHNCLADVPEVLGNLOKQIVPRVNVIDLKLGGLGCDIGL  
KGLIAPSKMRFNLYVNPYPCERIFGKRKENVYNKHPAGLWKLKPKPSNLTTKED  
VQNMKSFISSLA"  
2203..2222  
/rpt\_family="(A)n"  
4810..6511  
/gene="F2P9.2"  
/note="similar to cytochrome P-450 GB:AAB37231 from  
[Phalaenopsis sp. SM9108]"  
join(<4810..5811,5900..>6511)  
/gene="F2P9.2"  
join(4810..5811,5900..6511)  
/gene="F2P9.2"  
/codon\_start=1  
/product="putative cytochrome P-450; 4810-6511"  
/protein\_id="AAG52514.1"  
/db\_xref="GI:12325138"  
/translation="MTIDMYLSPASRSRGSPFFSLELCLSLFISLFVFWLTPGGFA  
WALYKARHTPESTQTGPAIPGSGPLIFGLLAFVNNALTHRLILANIADTCAKALM  
AFVSGSTFVITSPEETAKELLASSAFADSPKESAYELLFDRAMGPAFGDYRELK  
RISSTHLPSPKRISSGESRRKIQONVGEIKNAKECYGEVHLIKILHFGSANNVMS  
VGRKYNFEGIVISKESNEHLVSEGYELLGIDFVDSHFFPGMRWLDLOGVRRRCS  
LVGRVNVFGKIIINDHKKSLRDNPSESTYDDDFVLLMGHNSKLSLSDMLAVLW  
EMIFGTDVAILLEWLARWLHPDIOAKAABIDCIVGDSGRQVDSLPKLPYVR  
AIVKTLRMHPGPELLSARLSIHDTQIGTHFIPAGTAMVNMWALTHTDSKVPFAHE  
YKPERFLGQSNFPIMGSDRLAPFAGGRVCPGKMGCLATVLMWLAQLLSYKVV  
SCGEVDLSETLKLSEMKNTLVCAIPRG"  
complement(6762..6783)  
/rpt\_family="AT-rich"  
7673..8831  
/gene="F2P9.3"  
/note="similar to flavonol 4'-sulfotransferase GB:P52837  
from [Flaveria chloraefolia]"  
<7673..8831  
/gene="F2P9.3"  
7673..8689  
/gene="F2P9.3"  
/codon\_start=1  
/product="putative flavonol sulfotransferase; 7673-8689"  
/protein\_id="AAG52515.1"  
/db\_xref="GI:12325136"  
/translation="MESKTTQNGSEVVVELTEFEKTKQYQDFIATLPKSGWRPDEIL  
TOYQGWQECLEGLFHAKDFEARTDFVLCVSYPTGTTWALKATYAIWNSRYDD  
AANLKNRPHFVYVEIDFAFYPTVDVLQDRKNPLFSTHINGLPDSIVNSGCMQ  
VYMRDQDTFISMTWTLHREKSGSGLASLSDSDFMFCGLSVYGLDHLVLYKKA  
YOENRDLFLRYETMRANPLPFVKLAEMGYGFTDEERENGVAEKVKLCSFETLK  
NLEANKGKEREDRPVYVANSAYFRKKGVDWANYLTPMAARIDGLVEEKFKDTGLL  
QHDN"  
complement(9123..10175)  
/gene="F2P9.4"  
/note="similar to FLAVONOL 4'-SULFOTRANSFERASE GB:P52837  
from [Flaveria chloraefolia]"  
complement(<9123..>10175)  
/gene="F2P9.4"  
complement(9123..10175)  
/gene="F2P9.4"  
/codon\_start=1  
/product="putative flavonol sulfotransferase; 10175-9123"  
/protein\_id="AAG52515.1"  
/db\_xref="GI:12325139"  
/translation="MESLTLAKATITTTLPSSHDTKSTESTPEKQKQYQDLISTF  
PHEKGRPKPLIEVGYWMLPSLEGICHAQEFQARPSDFLVSYPKGTGTTWALK  
TFAIANRSRFDSSNLLKRNPHSEVYIEIDFPFVVDVLDKKGNTLPSTHLYPEL  
LPDSVYKSGCKMYIWRPKDTFTSMWTLHKETELGPVNSLEESFDMRCGLSGYV  
KYNNHILAYKWAYQENRDLFLFYETNRADPLFYKSLARFVGHGFTASEEKGVVR  
KYVNLCSFETLKNLEANKGKEREDRPGVYVANSAYFRKKGVDWNSYLTPEMAARIDG  
LMEKFKGTGLLEHGK"  
complement(12228..12256)  
/rpt\_family="AT-rich"  
complement(12462..12498)  
/rpt\_family="AT-rich"  
complement(12549..13136)  
repeat\_region



RESULT 12	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	PUBMED	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT	FEATURES	source	ORIGIN	Alignment Scores:	Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:
AY087493	1347 bp	mRNA	linear	PLN 14-APR-2003																						
Arabidopsis thaliana clone 36020 mRNA, complete sequence.																										
AY087493	1	GI:21406230																								
FLI CDNA.																										
Arabidopsis thaliana (thale cress)																										
Arabidopsis thaliana																										
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi																										
1 (bases 1 to 1347)																										
Haas,B.J., Volkovskiy,N., Town,C.D., Troukhan,M., Alexandrov,N., Feldmann,K.A., Flavell,R.B., White,O. and Salzberg,S.L.																										
Full-length messenger RNA sequences greatly improve genome annotation																										
Genome Biol. 3 (6), RESEARCH0029 (2002)																										
22088475																										
12093376																										
2 (bases 1 to 1347)																										
Brover,V., Troukhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and Feldmann,K.																										
Full-length cDNA from Arabidopsis thaliana																										
Unpublished																										
3 (bases 1 to 1347)																										
Brover,V., Troukhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and Feldmann,K.																										
Direct Submission																										
Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road, Malibu, CA 90265, USA																										
This clone sequence is one of 5,000 Ceres full-length cDNAs made available to TIGR and Genbank. The following quality assessment of this set was done by comparison with known proteins: two percent of the clones are estimated to be 5'-truncated; less than one percent are 3'-truncated; approximately two percent represent alternative splice variants, including unspliced introns and spliced exons; one percent may contain premature stop codons; five percent may have frame shifts in a coding region. A sequence is considered to be 5'-truncated if it lacks the translation initiation start (ATG). A C-terminal end of the encoded protein. Please note that these cDNA sequences are derived from the Ws or Laer ecotypes and therefore may contain polymorphisms when compared to sequences from Col-0. Genset carried out the library production and sequencing of the full-length clones. Ceres, Inc. carried out the clustering of the 5' sequences, selection of clones, and sequence assembly.																										
Location/Qualifiers																										
1..1347																										
/organism="Arabidopsis thaliana"																										
/mol_type="mRNA"																										
/db_xref="taxon:3702"																										
/clone="36020"																										
190..1206																										
/codon_start=1																										







```

Db      733 CAAGAGATCCAGATAGGATTTTGTCTTCTAGGTACGACCATGAGGCCAATCCCTTTG 792
QY      259 GluAsnValLysLeuAlaGluPheMetGlyCysGlyPheThrAspAspGluLys 278
Db      793 CCTTTTGTGAAGAGATTGCTGAGTTTCATGGGTTATGATTCACATGAGAGAGGAG 852
QY      279 GlnGlyIleValAspGluLysLeuValLysLeuValLysLeuValLysLeuValLys 298
Db      853 ATGGTGTGCTGAGAAAGTGTGAAGCTTTGAGCTTTGAGACCTTGAAGATCTTCAA 912
QY      299 ValAsnLysAsnGlySerSerTyrrAsnSerLys-----IleAspAsnLysHisPhe 315
Db      913 GCTAACAAAGGTCATAAAGAAAGAGAGGATCGTCTGCTGTTTATGCGAATAGCCGAT 972
QY      316 PheArgLysGlyGluValArgAspTrpAlaAsnTyrrLeuThrSerGluMetIleLysLys 335
Db      973 TTATGAAAGAAAGTGTGAGATTTGGGTAATATTATTCACCTCCTGAGATGCTCGT 1032
QY      336 LeuGluThrAlaGlyLysIleAsnGluSerGluLysHis-----LeuLeuSerAsnLys 353
Db      1033 ATTGAT-----GGCTTAGTGAGGAGAAATTCMAAGATACCTGGCTTGTCTCAACATGAT 1086
QY      354 AsnLeuThrAsnPheAsnLeuAsnAsnTyrrCys-----GluLysSer 367
Db      1087 AACTGAATGTGTTTATCTCGG--TCCTTCTGTATTGCTGCGGCGAAACAGAGGATGC 1144
QY      368 AsnIleSerLeuCysLeuLeuSerTyrrSer 377
Db      1145 TCTGTTTGTGTTTGTGTCAGTATCT 1174

RESULT 15
AC146342/c
LOCUS   Medicago truncatula clone mth2-10c20, WORKING DRAFT SEQUENCE, 2
DEFINITION
unordered pieces.
ACCESSION AC146342.4 GI:38488819
VERSION    HTG; HTGS PHASE1; HTGS DRAFT.
KEYWORDS   Medicago truncatula (barrel medic)
SOURCE     Medicago truncatula
ORGANISM   Medicago
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE  1 (bases 1 to 113531)
            Lin.S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
            and Roe,B.A.
            Medicago truncatula BAC Clone mth2-10c20
            Unpublished
REFERENCE  2 (bases 1 to 113531)
            Lin.S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
            and Roe,B.A.
            Direct Submission
            Submitted (12-AUG-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE  3 (bases 1 to 113531)
            Lin.S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
            and Roe,B.A.
            Direct Submission
            Submitted (21-NOV-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT   On Nov 21, 2003 this sequence version replaced gi:38371842.
            ----- Genome Center.
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR

```

\* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 27561: contig of 27561 bp in length  
 27562 27661: gap of unknown length  
 27662 113531: contig of 85870 bp in length.

FEATURES  
 source  
 1..113531  
 /organism="Medicago truncatula"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:3880"  
 /clone="mth2-10c20"  
 /clone\_lib="Medicago truncatula BAC library H2"

## ORIGIN

Alignment Scores:  
 Pred. No.: 6.87e-37 Length: 113531  
 Score: 567.50 Matches: 125  
 Percent Similarity: 57.27% Conservative: 64  
 Best Local Similarity: 37.88% Mismatches: 122  
 Query Match: 27.56% Indels: 19  
 DB: 2 Gaps: 10

US-09-854-122-16 (1-391) x AC146342 (1-113531)

```

QY      37 AspSerLysMetTyrrLysArgTyrrArgGluIleValSerSerLeuPro----SerAsnAsp 55
Db      109250 GAAACAAGGTAATCCATGATTATAAACAATATCTTGTCACTTCCTAGAGAAATGGG 109191
QY      56 TyrTrpGlyAspThrMetArgLeuTyrrLysGlyPheTrpGlnMetGlyTyrrLeuValPro 75
Db      109190 TGTGAGACTGAATATATGATATTTTTCATGGATTTTGGTGCCATCAATATCTTATCCAA 109131
QY      76 GlyIleMetAlaPheGluAspAsnPheLysAlaArgGluThrAspIleLeuThrThr 95
Db      109130 TCTGTCAACACTTTTCAAAATAAATTCATGCCAAGACAAATGATATTTGTTCATCT 109071
QY      96 LeuProLysAlaGlyThrThrTrpThrLysAlaLeuThrPheAlaIleLeuThrArgAsp 115
Db      109070 ATGCCAAATATGGCACCACTTGCTTAAAGGCTTCATATGCTATTGTATATCGC--- 109014
QY      116 ValAsnHisProSerSer-----ProThrHisProLeuLeuPhePheAsnProHis 132
Db      109013 ---CAACATTTTACCTCTCTTAGAGATAATAACCATCCATCTATTATTATATCCACAT 108957
QY      133 SerCysValGlnAsnLeuGluTyrr---LeuTyrrMetGlyArgGluAsnThrMetProAsp 151
Db      108956 GAACATAGTGCCTCTGTTTGAAGGTAAACCTTTATGGTGGTAAAGACATCTTTTGGCCTCAA 108897
QY      152 LeuAspMetLeuAsnGlu---SerProArgLeuPheAlaGlyHisIleProTyrrSerLeu 170
Db      108896 ATTGATGTATCAATATGATTGAACCAAGGCTTTTGGCACTCACATTCCTCTTCG 108837
QY      171 LeuProLysSerValLeuLysSerGlyThrLysIleIleAsnIleSerArgAsnArgLys 190
Db      108836 TTGCCCAAGTCGGTTAAAGAGTCAGTGTAAATCAATTTATATATGATGAACCCCAATTT 108777
QY      191 SerThrPheValSerPheTrpLysPheGlyAsnLeuIleAsnProAspLysLeu----- 208
Db      108776 GATACATTTGTATCATATTGGAATTATCAACAAGATCAGATCGAAGAATCTTTAACC 108717
QY      209 ---LeuAspLeuGluLysSerValAspIlePheAlaSerGlyIleSerPheCysGlyPro 227
Db      108716 GAATTAACCTTTAGAGGAGTCTTTTGAAGGTAATTTGTAAGGGAATATGTTTATTGTGGTCCG 108657
QY      228 GluTrp---AsnPheGlnAlaGluPheThrAsnAlaAlaSerThrAsnSerAsnLeuLeu 246
Db      108656 TTTTGGGACAAATATGTTGGGTACTTCTGAAGAGAGTATAGAAGACACACATAGGTTTAA 108597
QY      247 LeuLeuSerTyrrGluGluMetLeuGluLysProValGluAsnValLysLysLeuAlaGlu 266
Db      108596 TCTTTGAAATATGAAGACCTTAAAGAGATGTTAATTTTACATCAAAAGAAATTCAGAG 108537

```



Mon Mar 22 09:47:42 2004

us-09-854-122-16.rge

```

Qy 267 PheMetGlyCysGlyPheThrAspGluGluLysGlnGlyIleValAspGluIleVal 286
Db 108536 TTTGGGATTTCTTTCATCTCAAGAGAGAAATAATGGAGTGGTTGAAATATAATC 108477
Qy 287 LysLeuCysSerPheAspAsnLeuLysAsnGlnValAsnLysAsnGlySerSerTyr 306
Db 108476 AAGTTATGCAGCTTTCAGAGTATGAGGAATCAAGGAAATCAATCT--GAAACTGTA 108420
Qy 307 AsnSerLysIleAspAsnLysHisPheArgLysGlyGluValArgAspTyrPalaAsn 326
Db 108419 TCACAAATCTATGAGAAAGTTTCTTTCGAAAGGAGATAGAGATTGGGTAAT 108360
Qy 327 TyrLeuThrSerGluMetIleLysLysLeuGluThrAlaGlyLysIleAsnGluSerGlu 346
Db 108359 TACTTTCTCTCTTCAATGATAGAAAGATTA-----TCCAAAGTCATAGAGAAAAA 108309
Qy 347 LysHisLeuSerAsnLysAsnLeuThr 356
Db 108308 -----TTAGTGGATCAACCTATCA 108288

```

Search completed: March 20, 2004, 05:31:28  
Job time : 5431 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 20, 2004, 01:48:09 ; Search time 507 Seconds

(without alignments)  
3276.224 Million cell updates/sec

Title: US-09-854-122-16

Perfect score: 2059  
Sequence: 1 TRGTGTGIAVACLPLMAGIL.....CLLSYFVNNFILLKKKK 391

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame\_plus\_p2n.model -DEV=xlh  
-Q=/cgr2\_1/USPTO.spool/US09854122/runat\_18032004\_100000\_3247/app\_query.fasta\_1.583  
-DB=N Geneseq 29Jan04 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09854122.CGN\_1\_1\_470 @runat\_18032004\_100000\_3247 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 29Jan04:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002as:\*  
7: Geneseqn2003as:\*  
8: Geneseqn2003bs:\*  
9: Geneseqn2003cs:\*  
10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	1996	96.9	1192	6 AAS20863	Aas20863 Z. marina
2	573	27.8	1347	3 Aac37555	Aac37555 Arabidops
3	572.5	27.8	1212	3 Aac39344	Aac39344 Arabidops
4	567	27.5	1041	7 Abz68953	Abz68953 Nucleotid
5	566	27.5	1281	3 Aac37716	Aac37716 Arabidops
6	562.5	27.3	1017	7 Abz68951	Abz68951 Nucleotid
7	558.5	27.1	1053	7 Abz68952	Abz68952 Nucleotid
8	558	27.1	1400	7 Aad54410	Aad54410 Lolium pe

9	557.5	27.1	1205	3 AAC37521	Aac37521 Arabidops
10	555.5	27.0	1077	4 Aaf29177	Aaf29177 HydroxyJa
11	555.5	27.0	1270	3 Aac48839	Aac48839 Arabidops
12	555.5	27.0	1273	3 Aac47848	Aac47848 Arabidops
13	548.5	26.6	1154	3 Aac37672	Aac37672 Arabidops
14	545	26.5	936	6 ABZ13222	Abz13222 Arabidops
15	537.5	26.1	980	7 ADA68055	Ada68055 Arabidops
16	537.5	26.1	981	6 ABZ14244	Abz14244 Arabidops
17	537.5	26.1	981	7 ABZ42036	Abz42036 Arabidops
18	537.5	26.1	1043	3 AAC42382	Aac42382 Arabidops
19	528.5	25.7	999	7 ADA69659	Ada69659 Rice gene
20	527	25.6	271990	9 ADD25213	Add25213 Fertility
21	521.5	25.3	1017	7 ADA69350	Ada69350 Rice gene
22	515.5	25.0	1160	3 AAC32748	Aac32748 Arabidops
23	510.5	24.8	1041	4 AAF29178	Aaf29178 HydroxyJa
24	483	23.5	2067	9 ADD25206	Add25206 Fertility
25	479	23.3	1032	7 ADA71112	Ada71112 Rice gene
26	446.5	21.7	1040	7 ADA71367	Ada71367 Rice gene
27	443.5	21.5	1460	7 ADA70067	Ada70067 Rice gene
28	409	19.9	776	7 Aad54404	Aad54404 Lolium pe
29	396	19.2	776	7 Aad54402	Aad54402 Lolium pe
30	372	18.1	777	7 Aad54405	Aad54405 Lolium pe
31	338.5	16.4	1000	8 ADB23209	Adb23209 Environme
32	330	16.0	2000	7 ADA71733	Ada71733 Rice gene
33	321	15.6	515	9 ADE81349	Ade81349 Arabidops
34	317.5	15.4	668	7 Aad54403	Aad54403 Lolium pe
35	273	13.3	1063	2 AAV11154	Aav11154 Human EST
36	273	13.3	1063	2 AAV12466	Aav12466 Human oes
37	270.5	13.1	566	7 ABX56987	Abx56987 Arabidops
38	268	13.0	1004	4 AAD19778	Aad19778 Human ST
39	268	13.0	1240	4 AAD09944	Aad09944 Human dru
40	264	12.8	942	4 AAD19776	Aad19776 Human ST
41	262	12.7	1600	6 ABN83170	Abn83170 Human sul
42	260	12.6	915	6 ABN83168	Abn83168 Human sul
43	260	12.6	922	4 AAD19777	Aad19777 Human ST
44	259.5	12.6	849	6 ABN83171	Abn83171 Human sul
45	257	12.5	1363	6 ABK63484	Abk63484 Rat seque

#### ALIGNMENTS

RESULT 1

AAS20863  
ID AAS20863 standard; cDNA; 1192 BP.

XX AAS20863;

XX 09-APR-2002 (first entry)

DE Z. marina cDNA clone encoding protein containing sulfotransferase.

XX Plant; transgenic; marine eelgrass; zosteric acid biosynthesis;  
XX saline-resistance; anoxia-resistance; anti-fouling genetic trait;  
XX marine vascular plant; sulphated phenolic compound; Zostera marina;  
XX sulfotransferase; ST; enzyme; gene; ss.

XX Zostera marina.

XX Key

FT Location/Qualifiers

FT 1..1192

FT /\*tag= a

FT /partial

FT /product= "Protein containing sulfotransferase"

FT /note= "This sequence lacks both start and stop codons"

FT /transl\_except= (pos:34..36, aa:Xaa)

FT /transl\_except= (pos:1042..1044, aa:Xaa)

FT /transl\_except= (pos:1060..1062, aa:Xaa)

FT /transl\_except= (pos:1078..1080, aa:Xaa)

FT /transl\_except= (pos:1087..1089, aa:Xaa)

FT /transl\_except= (pos:1120..1122, aa:Xaa)

FT /note= "Xaa= Stop codon"

XX WO200185971-A2.

XX 15-NOV-2001.  
PD 10-MAY-2001; 2001WO-US015412.  
XX PF 10-MAY-2000; 2000US-0202529P.  
XX PR (PHYC-) PHYCOGEN INC.  
XX PA Alberte RS, Smith RD;  
XX PI WPI; 2002-121947/16.  
XX DR P-PSDB; AAU12056.  
XX PT New transgenic plants comprising a zosteric acid biosynthetic gene, a  
PT saline resistance gene or a hypoxia resistance gene derived from *Zostera*  
PT marina, useful for producing plants with antifouling traits.  
XX Example; Fig 4; 117pp; English.  
XX PS The present invention relates to a new transgenic plant comprising a  
CC heterologous gene derived from the marine eelgrass *Zostera marina* or at  
CC least one heterologous nucleotide sequence encoding a zosteric acid  
CC biosynthetic function, a saline-resistance function, or a anoxia-  
CC resistance function. The invention describes the method of producing a  
CC transgenic plant possessing an anti-fouling genetic trait by providing a  
CC cDNA population derived from a marine vascular plant, isolating from the  
CC cDNA population a nucleic acid species which hybridizes to a nucleic acid  
CC that encodes a sulfotransferase (ST), an alcohol dehydrogenase (ADH), and  
CC phenylalanine ammonia lyase (PAL) or a cinnamate-4-hydroxylase (CH), and  
CC transforming a target host plant with the isolated nucleic acid. The  
CC plant is useful in the genetic engineering of plant species having  
CC desirable genetic traits such as antifouling traits, salt and anoxia  
CC resistance, and pathogen defence strategy. The expression of such  
CC biosynthetic enzymes are sufficient to support the production of zosteric  
CC acid and other sulphated phenolic compounds in a target plant. The  
CC present sequence represents a *Z. marina* cDNA clone which encodes a  
CC protein containing sulfotransferase  
XX SQ Sequence 1192 BP; 391 A; 206 C; 266 G; 329 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 2,32e-185 Length: 1192  
Score: 1995.00 Matches: 391  
Percent Similarity: 98.49% Conservative: 0  
Best Local Similarity: 98.49% Mismatches: 0  
Query Match: 96.94% Indels: 6  
DB: 6 Gaps: 6

US-09-854-122-16 (1-391) x AAS20863 (1-1192)

QY 1 ThrArgGlyIleThrGlyIleAlaValAlaCys---LeuProLeuIleMetAlaGlyIle 19  
DB 1 AGCGGGGAATAACTGGGAATCGCTGTTGCTTGTAGCTACCACCTGTAATGGCTGGAAAT 60  
QY 20 LeuAlaLeuGluLysCysPheGlySerLysAsnGluGlnGluLysGluGluAspSerLys 39  
DB 61 TTATGCTTTGGGAATATGTTCCGATCCGAATGAGCAAGAGAGAGAGAGATTCACAA 120  
QY 40 MetTyrLysArgTyrArgGluIleValSerSerLeuProSerAsnAspTyrTyrGlyAsp 59  
DB 121 ATGTACAAGAGATATAGAGAGATGTTTCTTCACTTCCCTCCCTCGAATGATATGGGGGAT 180  
QY 60 ThrMetArgLeuTyrLysGlyPheTrpGlnMetGlyTyrLeuValProGlyIleMetAla 79  
DB 181 ACCATGAGGTTGTACAGGGATTTGGCAATGGGATATCTGTACCTGGTATCATGGCT 240  
QY 80 PheGluAspAsnPhelysAlaArgGluThrAspIleIleLeuThrThrLeuProLysAla 99  
DB 241 TTCGAAGATAATTTCAAGCTCGAGACGCGACATTATCTTACACCTCTTCCAAAGCT 300  
QY 100 GlyThrThrTyrThrLysAlaLeuThrPheAlaIleLeuThrArgAspValAsnHisPro 119

DB 301 GGAACGACATGGACGAAGGCACTGACGTTTGCCATCTAAACACGAGATGTTAACCCACCA 360  
QY 120 SerSerProThrHisProLeuLeuPhePheAsnProHisSerCysValGlnAsnLeuGlu 139  
DB 361 TCATCACCGACACATCCACTTTTGTCTTCAACCTCATTCGTGTGTTCAAAATTTGAG 420  
QY 140 TyrLeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsnGluSerPro 159  
DB 421 TATTTGTACATGGGTAGAGAAATACGATGCCACCTCGATATGTTGAATGAATCGCG 480  
QY 160 ArgLeuPheAlaGlyHisIleProTyrSerLeuLeuProAlaSerValLeuLysSerGly 179  
DB 481 AGGTTGTTTGGCGACACATCCCATCTTTTGTGCGGGCTCTGTTTGAATTCGGGA 540  
QY 180 ThrLysIleIleAsnIleSerArgAsnArgLysSerThrPheValSerPheTyrLysPhe 199  
DB 541 ACAAAATCATCAATATAAGCCGACACCGTAAGAGTACATTTGTCTTTTGGAAATTT 600  
QY 200 GlyAsnLeuIleAsnProAspLysLeuAspLeuLysSerValAspIlePheAla 219  
DB 601 GGCATCTGATTAAACCCGACAGTTATTGGACCTCGAAAGAGCGTTGATATCTTCGCA 660  
QY 220 SerGlyIleSerPheCysGlyProGluTyrAsnPheGlnAlaGluPheThrAsnAlaAla 239  
DB 661 TCGGGAATCTCTCTTTTGTGGACCGAATGGAATTTCCAGCGGAGTTCCACCAATCGCG 720  
QY 240 SerThrAsnSerAsnLeuLeuLeuSerTyrGluGluMetLeuGluLysProValGlu 259  
DB 721 TCTACTAATTCAAACTTCTGTTTGTAGTTTACGAAGAAATGTTAGAGAAGCCAGTTGAA 780  
QY 260 AsnValLysLysLeuAlaGluPheMetGlyCysGlyPheThrAspAspGluLysGln 279  
DB 781 AATGTGAAAGAGCTAGCTAGTTCTATGGGATGTGGTTTCAGACGATGAGGAGAAACAA 840  
QY 280 GlyIleValAspGluIleValLysLeuCysSerPheAspAsnLeuLysAsnGlnVal 299  
DB 841 GGGATTTGTGATGAGATAGTTAAACTTTTGTAGCTTCGACAAATCTGAAGAATCAACAG 900  
QY 300 AsnLysAsnGlySerSerTyrAsnSerLysIleAspAsnLysHisPheArgLysGly 319  
DB 901 AACAAACCGGATCAAGCTCAATTCGAATTCGACACACAGCATTTCTTCAGGAAGGT 960  
QY 320 GluValArgAspTyrAlaAsnTyrLeuThrSerGluMetIleLysLysLeuGluThrAla 339  
DB 961 GAGGTGAGAGATTTGGCAAACTATCTAACTGCGGAATGATTAAGAACTGGAGACGGCC 1020  
QY 340 GlyLysIleAsnGluSerGlu---LysHisLeuLeuSer---AsnLysAsnLeuThr--- 356  
DB 1021 GGAAAAATAAATGAATCAGAGTAAAGCATTTTATTTCTGTGAATAGAAATCTTACATGA 1080  
QY 357 AsnPhe---AsnLeuAsnAsnTyrCysGluLysSerAsn---IleSerLeuCysLeuLeu 374  
DB 1081 AACTTCTGAAATCTTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1140  
QY 375 SerTyrSerPheValIleAsnAsnPhelIleLeuLeuLysLysLysLys 391  
DB 1141 TCGTATTCATTCGTATAATAAATTAATTTCTATTGTTTAAAAAATAAATAAATAA 1191

RESULT 2  
AAC37555  
ID AAC37555 standard; DNA; 1347 BP.  
XX AAC37555;  
XX AAC37555;  
DT 17-OCT-2000 (first entry)  
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 17816.  
XX Hybridisation assay; genetic mapping; gene expression control;  
XX protein identification; signal transduction pathway; metabolic pathway;  
XX promoter; termination sequence; ss.  
OS Arabidopsis thaliana.

XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-00301439.  
XX XX 99US-0121825P.  
PR 05-FEB-1999; 99US-0123180P.  
PR 05-MAR-1999; 99US-0123348P.  
PR 03-MAR-1999; 99US-0125788P.  
PR 23-MAR-1999; 99US-0126264P.  
PR 23-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131149P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139453P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141267P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142330P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144351P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145152P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148174P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 16-AUG-1999; 99US-0148684P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149468P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.



KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway; metabolic pathway;  
XX promoter; termination sequence; ss.  
OS Arabidopsis thaliana.  
PN EP1033405-A2.  
XX 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-00301439.  
XX  
FF 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 28-APR-1999; 99US-0130891P.  
PR 30-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 04-MAY-1999; 99US-0132407P.  
PR 05-MAY-1999; 99US-0132484P.  
PR 06-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 07-MAY-1999; 99US-0132487P.  
PR 11-MAY-1999; 99US-0132863P.  
PR 14-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 21-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139859P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145122P.  
PR 22-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147132P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149486P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.



```

FT FT /*tag= a
XX FT /product= "desulfoglucosinolate sulfotransferase AtST5c"
XX PN
XX PN WO2003010318-A2.
XX PD 06-FEB-2003.
XX XX
XX PD 24-JUL-2002; 2002WO-CA001144.
XX PF
XX XX
XX FR 24-JUL-2001; 2001US-0307141P.
XX XX
XX XX (UYCO-) UNIV CONCORDIA.
XX XX
XX PI Varin L, Spertini D;
XX XX
XX DR WPI; 2003-248082/24.
XX DR F-PSDB; ABP72364.
XX XX
XX PT Novel isolated or purified polypeptide having biological activity of
XX PT desulfoglucosinolate sulfotransferases, useful as enzymes involved in
XX PT glucosinolate biosynthesis in transgenic plants.
XX PS
XX PS Claim 9; Page 60-62; 68pp; English.
XX CC
XX CC The present sequence encodes a plant desulfoglucosinolate
XX CC sulfotransferase, designated AtST5c. The specification also describes
XX CC AtST5a and AtST5b. AtST5a and AtST5b are constitutively expressed in all
XX CC parts of the plants, while AtST5c does not seem to be expressed in the
XX CC early stage of development. These enzymes are involved in glucosinolate
XX CC synthesis in plants. The desulfoglucosinolate sulfotransferase
XX CC polynucleotides are useful for modulating glucosinolate biosynthesis in
XX CC transgenic plants
XX XX
XX SQ Sequence 1041 BP; 303 A; 208 C; 240 G; 290 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 1e-45 Length: 1041
Score: 587.00 Matches: 137
Percent Similarity: 55.78% Conservations: 56
Best Local Similarity: 39.60% Mismatches: 123
Query Match: 27.54% Indels: 30
DB: Gaps: 11

US-09-854-122-16 (1-391) x ABZ68953 (1-1041)

Qy 23 GluLysCysPheGlySerLysAsnGlnGlnGlyGluGluAspSerLysMetTyrLys 42
Db 49 GAGTTAGCATCTCGCGCCATCAAAATTCAGAAATTCAGAGAACGAG ---AAA 93
Qy 43 ArgTyrArgGluIleValSerSerLeuProSerAsnAspTyrTrp-----GlyAspThr 60
Db 94 CATTACCAAGAAATCATCGCAACTCTTCTCACAAAGATGGCTGAGACCCAAAAGATCGG 153
Qy 61 MetArgLeuTyrLysGlyPheTrpGlnMetGlyTyrIleValProGlyIleMetAlaPhe 80
Db 154 TTCGTTGAGTACGGTGGTCACTGGTGGGTGACAACTCTTTTGAAGGTTTACTTCACGGC 213
Qy 81 GluAspAsnPhelLysAlaArgGluThrAspIleLeuThrThrLeuProLysAlaGly 100
Db 214 CAGAAGTTCTTCAAGGCACGTCCTCCCAATGATTTTCGTCGCAGCTACCCAAAACCGGC 273
Qy 101 ThrThrTrpThrLysAlaLeuThrPheAlaIleLeuThrArg-----AspValAsn 117
Db 274 ACGACTTGGCTCAAGCCCTTAACCTTCGCAATTCGCAAAATCGCTCCAAAGTTCGACGTTTCA 333
Qy 118 HisProSerSerProThrHisProLeuLeuPhePheAsnProHisSerCysValGlnAsn 137
Db 334 -----ACAAACCTCTTCTCAAAACGCTATCCACAGGAGTTTGTGCCTTAC 378
Qy 138 LeuGluTyrLeuTyrMetGlyArgGluAenthMetProAspLeuLeuAspMetLeu---Asn 156
Db 379 ATCGAAATCGACTTC-----CGTTTTTCCCAAGCGTTGATGTTCTTTAAGAC 476

```



PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130810P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132240P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-01334256P.  
PR 14-MAY-1999; 99US-01334218P.  
PR 14-MAY-1999; 99US-01334219P.  
PR 14-MAY-1999; 99US-01334221P.  
PR 14-MAY-1999; 99US-01334370P.  
PR 18-MAY-1999; 99US-01334768P.  
PR 19-MAY-1999; 99US-01334941P.  
PR 20-MAY-1999; 99US-0133524P.  
PR 21-MAY-1999; 99US-01335353P.  
PR 24-MAY-1999; 99US-01335629P.  
PR 25-MAY-1999; 99US-01336021P.  
PR 27-MAY-1999; 99US-01336392P.  
PR 28-MAY-1999; 99US-01336782P.  
PR 01-JUN-1999; 99US-01337232P.  
PR 03-JUN-1999; 99US-01337528P.  
PR 04-JUN-1999; 99US-01337502P.  
PR 07-JUN-1999; 99US-01337724P.  
PR 08-JUN-1999; 99US-01338094P.  
PR 10-JUN-1999; 99US-01338540P.  
PR 10-JUN-1999; 99US-01338847P.  
PR 14-JUN-1999; 99US-01339119P.  
PR 16-JUN-1999; 99US-01339452P.  
PR 16-JUN-1999; 99US-01339453P.  
PR 17-JUN-1999; 99US-01339432P.  
PR 18-JUN-1999; 99US-01339454P.  
PR 18-JUN-1999; 99US-01339455P.  
PR 18-JUN-1999; 99US-01339456P.  
PR 18-JUN-1999; 99US-01339457P.  
PR 18-JUN-1999; 99US-01339458P.  
PR 18-JUN-1999; 99US-01339459P.  
PR 18-JUN-1999; 99US-01339460P.  
PR 18-JUN-1999; 99US-01339461P.  
PR 18-JUN-1999; 99US-01339462P.  
PR 18-JUN-1999; 99US-01339463P.  
PR 18-JUN-1999; 99US-01339750P.  
PR 18-JUN-1999; 99US-01339763P.  
PR 21-JUN-1999; 99US-01339817P.  
PR 22-JUN-1999; 99US-01339899P.  
PR 23-JUN-1999; 99US-0140333P.  
PR 24-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 19-JUL-1999; 99US-0144336P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144684P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147303P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 05-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 06-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 13-AUG-1999; 99US-0148685P.  
PR 18-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150684P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 30-AUG-1999; 99US-0151080P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0158458P.  
PR 29-SEP-1999; 99US-0158598P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0159369P.  
PR 13-OCT-1999; 99US-0159293P.

```

PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 25-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:
Pred. No.: 1,66e-45 Length: 1281
Score: 566.00 Matches: 143
Percent Similarity: 54.31% Conservative: 65
Best Local Similarity: 37.34% Mismatches: 139
Query Match: 27.49% Indels: 36
DB: 3 Gaps: 12

US-09-854-122-16 (1-391) x AAC37716 (1-1281)

QY 29 LysAsnGluInGluLysGluLysAspSerLysMetTyrLysArgTyrArgGluLeuVal 48
DB 159 GAGACCAAGACAGAAATCAACAGAGTTCAGAGAAATCAAAACCGTATCAAGACCTCATC 218
QY 49 SerSerLeuProSerAsnAspTyrTrpGlyAspThrMetArgLeu-----TyrLysGly 66
DB 219 TCCAGCTTCTCCAGAGAAAGCGTGGAGCCGAAAGAGCCCTCGATCGAGTATGGTGGT 278
QY 67 PheTrpGlnMetGlyTyrLeuValProGlyLeuMetAlaPheGluAspAsnLysAla 86
DB 279 TACGTGGTACCGTCTCTCTCGAAGGTGTATTCAAGCGCAAGAGTCTTTCAAGCA 338
QY 87 ArgGluThrAspIleLeuLeuThrThrLeuProLysAlaGlyThrThrTrpThrLysAla 106
DB 339 CGACCCAGTACTCTCTCGTGTAGCTTACCCAAAGACAGCACCACTTGGCTCAAAGCC 398
QY 107 LeuThrPheAlaIleLeuThrArgAspValAsnHisProSerSerProThrHisProLeu 126
DB 399 CTGACTTTCCGCAATCGTTCCTCGCTTCGATGATTCCTCC-----AACCTCTC 452
QY 127 LeuPhePheAsnProHisSerCysValGlnAsnLeuGluTyrLeuTyrMetGlyArgGlu 146
DB 453 CTGAACAGTAACCTCTCAGAGTTGTTCCTTACATTGAGATAGATTTC----- 500
QY 147 AsnThrMetProAspLeuAspMetLeuAsnGluSerProArg---LeuPheAlaGlyHis 165
DB 501 CCTTCTTCTCCCTGAAGTTGATGTTCTCAAGACAAAGGGAACACTCTGTTTCGACTCAT 560
QY 166 IleProTyrSerLeuLeuProAlaSerValLeuLysSerGlyThrLysIleIleAsnIle 185
DB 561 ATCCCATACGAGTATTACCTGATTGCGGTTGTGAAATCCGGTTGTAAAGATGGTTTACATA 620
QY 186 SerArgAsnArgLysSerThrPheValSerPheThrLysPheGlyAsnLeuIleAsnPro 205
DB 621 TGGAGAGAACCAAGACACTTTTATCTCCATGCGACTTTC-----CTTCACAAG 671

206 AspLys-----LeuLeuAspLeuGluLysSerValAspIlePheAlaSer 220
672 GAAGACAGACAGCTTGGACCTTGCAGCAATCTTGGAGAGTCTTTGATATATTTCTGCTGT 731
221 GlyIleSerPheCysGlyProGluTrpAsnPheGlnAlaGluPheThrAsnAlaAlaSer 240
732 GGTCTGTCTGGGTATGTCCTTATCTTAACTATCATCTCTGGCGTATTGGAAAGCATACCAA 791
241 ThrAsn-----SerAsnLeuLeuLeuSerTyrGluGluMetLeuGluLysProValGlu 259
792 GAGAAATCCAGATAGATCTTCTTCAAGTACGAGAGATGAGAGCTGATCTTTCACCG 851
260 AsnValLysLysLeuAlaGluPheMetGlyCysGlyPheThrAspAspGluGluLysGln 279
852 TACGTGAAGAGTCTGGCTGAGTATTATGGTTCATGGAITTCACAGCGAGGAGAGAGAAA 911
280 GlyIleValAspGluIleValLysLeuCysSerPheAspAsnLeuLysAsnGlnVal 299
912 GGTGTGTGTGAAAGTGTGTAATCTTTCGAGCTTCGAGACGTTGAAGAAATCTTGAAGCT 971
300 AsnLysAsnGlySerSerTyrAsnSerLys-----IleAspAsnLysHisPhePhe 316
972 AACAAAGGGGAGAAAGACAGAGAGATCGTCTGTGTTTACGCAATAGCGGTATTTTC 1031
317 ArgLysGlyGluValArgAspTrpAlaAsnTyrIleuThrSerGluMetIleLysLysLeu 336
1032 AGGAAAGGAAAGCTGGGAGATTGTTGCACTATCTGACTCCGAGATGGCTGCTGCTATA 1091
337 GluThrAlaGlyLysIleAsnGluSerGluLysHis-----LeuLeuSerAsnLys 353
1092 GAT-----GGTTAATGGAGAGAAATTTAAGGCACCGCTTCTTGAACATGTTAAA 1145
354 AsnLeuThrAsnPheAsn-----LeuAsnAsnTyrCysGluLysSerAsnIle 369
1146 TGAATGACTTAATCTTGTTCGAGCAATGACTGATCTTTATATGCTTTTGTATTACACATA 1205
370 SerLeuCysLeuLeu-----SertYrSerPheValIleAsnAsnPhe 383
1206 ACTTTGCTTCTTATGTTATGATGCTCCAAAATGATTATTAAGTTATTATAGTGGAGAGATAA 1265
384 IleLeuLeu 386
1266 ATGTTGCTT 1274

RESULT 6
ABZ68951
ID ABZ68951 standard; DNA; 1017 BP.
XX AC ABZ68951;
XX DT 28-MAY-2003 (first entry)
XX DE Nucleotide sequence of desulfooglucosinolate sulfotransferase AtST5a.
XX KW Plant; desulfooglucosinolate sulfotransferase; AtST5a; AtST5b; AtST5c;
XX KM glucosinolate; transgenic plant; gene; ss.
XX OS Arabidopsis thaliana.
XX FH Key Location/Qualifiers
XX CDS 1..1017
FT FT /*tag= a
FT FT /product= "desulfooglucosinolate sulfotransferase AtST5a"
XX PN WC2003010318-A2.
XX PD 06-FEB-2003.
XX PF 24-JUL-2002; 2002WO-CA001144.
XX PR 24-JUL-2001; 2001US-0307141P.
XX

```

(UYCO-) UNIV CONCORDIA.  
 Varin L, Spertini D;  
 WPI; 2003-248082/24.  
 P-PSDB; ABP72962.  
 Novel isolated or purified polypeptide having biological activity of  
 desulfoglucosinolate sulfotransferases, useful as enzymes involved in  
 glucosinolate biosynthesis in transgenic plants.  
 Claim 9; Page 54-55; 68pp; English.  
 The present sequence encodes a plant desulfoglucosinolate  
 sulfotransferase, designated AtST5a. The specification also describe  
 AtST5b and AtST5c. AtST5a and AtST5b are constitutively expressed in all  
 parts of the plants, while AtST5c does not seem to be expressed in the  
 early stage of development. These enzymes are involved in glucosinolate  
 synthesis in plants. The desulfoglucosinolate sulfotransferase  
 polynucleotides are useful for modulating glucosinolate biosynthesis in  
 transgenic plants  
 Sequence 1017 BP; 296 A; 205 C; 240 G; 276 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 2,68e-45 Length: 1017  
 Score: 562.50 Matches: 129  
 Percent Similarity: 55.95% Conservative: 59  
 Best Local Similarity: 38.39% Mismatches: 123  
 Query Match: 27.32% Indels: 25  
 DB: 7 Gaps: 10

US-09-854-122-16 (1-391) x ABZ68951 (1-1017)

QY 33 GluLysGluLysSerLysMetLysArgLysGluLysValSerSerLeuPro 52  
 Db 40 GAACCTCAGAGCTTCGAGAAACCCAGAGAGAGTATCAAGATTTCATCGTACACTTCCA 99  
 QY 53 SerAsnAspTyrTrp-----GlyAspThrMetArgLeuTyrLysGlyPheTrpGlnMet 70  
 Db 100 AGAGCAAGGCTGGAGACCATGAGATCAACCAATACGGTGGACACTGGTGGCAA 159  
 QY 71 GlyTyrLeuValProGlyLysMetAlaPheGluAspAsnPheLysAlaArgLysThrAsp 90  
 Db 160 GAATGTCTCCTCGAAGTCTTTTTCACGCTAAAGACCAATTCGAAGCAGCAGCCAACTGAT 219  
 QY 91 IleIleLeuThrLeuProLysAlaGlyThrThrLysLysAlaLeuThrPheAla 110  
 Db 220 TTCCTCGTGTGATGATCCCAAAACCGGTACAACTTGGCTCAAGCACTACTTACCGA 279  
 QY 111 IleLeuThrArgAspValAsnHisProSerProThrHisProLeuLeuPhePheAsn 130  
 Db 280 ATCGTCAATCGT-----TCTCGTTACGACGACGCGCAACCCACTCTCTCAAGCAAAAC 333  
 QY 131 ProHisSerCysValGlnAsnLeuGluTyrLeuTyrMetGlyArgGluAsnThrMetPro 150  
 Db 334 CTTCAAGATTTTCCCTTACGTTAGATGATCGACTTCGG-----TTTTACCCC 381  
 QY 151 AspLeuAspMetLeuAsnGluSerProArg---LeuPheAlaGlyHisIleProTyrSer 169  
 Db 382 ACCGTTGATGTTCTTCAAGACAGAAAGAACCCACTTTTCTACTCATATCCCAACCGG 441  
 QY 170 LeuLeuProAlaSerValLeuLysSerGlyThrLysIleIleAsnLysSerArgAsnArg 189  
 Db 442 TTATTACCGGATTCGATTGGAATCTGGTTGTAAGATGGTGATCATATGAGAGACCGG 501  
 QY 190 LysSerThrPheValSerPheTrpLysPheGlyAsnLeuIleAsnProAspLys----- 207  
 Db 502 AAGATACITTCATCTCCATGTCGACTTC-----TTACACAGGAGAGAGTCTCAA 552  
 QY 208 -----LeuLeuAspLeuGluLysSerValAsnIlePheAlaSerGlyIleSerPhe.224  
 Db 553 GAAGGTCAATTAGCGAGTCTTGAGGACAGCTTTGATATGTTTGTAAAGCTTATCTGNG 612

QY 225 CysGlyProGluTyrAsnPheGlnAlaGluPheThrAsnAlaIleSerThrAsn---Ser 243  
 Db 613 TATGCTCTTATCTGATCATGTTTGGGTATTGGAAAGCTTACCAAGAGATCCAGAT 672  
 QY 244 AsnLeuLeuLeuSerTyrGluGluMetLeuGluLysProValGluAsnValLysLys 263  
 Db 673 AGGATTTTGTCTTAGGTACGAGACCATGAGGCGCAATCTTGGCTTTTGTGAAGAGA 732  
 QY 264 LeuAlaGluPheMetGlyCysGlyPheThrAspAspGluGluLysGlnGlyIleValAsp 283  
 Db 733 TTGGCTGATTTTCATGGGTTATGGATTCTGATGAGGAAGAGGAGAAATGGTGTCTGAG 792  
 QY 284 GluIleValLysLeuCysSerPheAspAsnLeuLysAsnGlnValAsnLysAsnGly 303  
 Db 793 AAGTGTGAGCTTTTGTAGCTTTGAGACGTTGAGATCTTGAAGCTAACAAAGGTGAT 852  
 QY 304 SerSerTyrAsnSerLys-----IleAspAsnLysHisPhePheArgLysGlyGlu 320  
 Db 853 AAAGAAGAGAGAGATGCTCTCTCTGTTTATCGAATACGGCGGTATTTTGAAGAAAGAAG 912  
 QY 321 ValArgAspTrpAlaAsnTyrLeuThrSerGluMetIleLysLysLeuGluThrAlaGly 340  
 Db 913 GTTGGAGATTGGCTAAATTATTTGACTCTCGAGATGGCTGCTCGATTGAT-----GGC 966  
 QY 341 LysIleAsnGluSerGluLysHis-----LeuLeuSerAsnLysAsn 354  
 Db 967 TTAGTGGAGAGAAATTCAGAGATACCTGGCTTGTCTCAACATGATAAC 1014  
 RESULT 7  
 ABZ68952  
 ID ABZ68952 standard; DNA; 1053 BP.  
 XX ABZ68952;  
 XX  
 XX 28-MAY-2003 (first entry)  
 DT  
 XX  
 DE Nucleotide sequence of desulfoglucosinolate sulfotransferase AtST5b.  
 XX  
 KW plant; desulfoglucosinolate sulfotransferase; AtST5a; AtST5b; AtST5c;  
 KW glucosinolate; transgenic plant; gene; ss.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 FH Key Location/Qualifiers  
 CDS 1..1053  
 FT /tag= a  
 FT /product= "desulfoglucosinolate sulfotransferase AtST5b"  
 FT  
 FN W02003010318-A2.  
 XX  
 PD 06-FEB-2003.  
 XX  
 PF 24-JUL-2002; 2002WO-CA001144.  
 XX  
 PR 24-JUL-2001; 2001US-0307141P.  
 XX  
 PA (UYCO-) UNIV CONCORDIA.  
 XX  
 XX Varin L, Spertini D;  
 XX WPI; 2003-248082/24.  
 XX P-PSDB; ABP72963.  
 DR  
 PT Novel isolated or purified polypeptide having biological activity of  
 PT desulfoglucosinolate sulfotransferases, useful as enzymes involved in  
 PT glucosinolate biosynthesis in transgenic plants.  
 XX  
 XX Claim 9; Page 57-59; 68pp; English.  
 XX  
 CC The present sequence encodes a plant desulfoglucosinolate  
 CC sulfotransferase, designated AtST5b. The specification also describes  
 CC AtST5a and AtST5c. AtST5a and AtST5b are constitutively expressed in all

parts of the plants, while AtST5c does not seem to be expressed in the early stage of development. These enzymes are involved in glucosinolate synthesis in plants. The desulfoglucosinolate sulfoxidase polynucleotides are useful for modulating glucosinolate biosynthesis in transgenic plants

XX  
SQ Sequence 1053 BP; 292 A; 235 C; 259 G; 267 T; 0 U; 0 Other;

Alignment Scores:	
Pred. No.:	6.9e-45
Score:	558.50
Length:	1053
Matches:	129
Percent Similarity:	59
Best Local Similarity:	Conservative:
Query Match:	38.9%
DB:	Mismatches: 120
	Indels: 23
	Gaps: 9

US-09-854-122-16 (1-391) X ABZ68952 (1-1053)

29 LysAsnGluGlnGluLysGluAspSerLysMetTyrLysArgTyrArgGluLeuVal 48  
:: ::  
64 GAGACCAGACAGAAATCAACAGAGTTCAGAGAAAATCAAAAACGGTATCAAGACCTCATC 123  
QY  
49 SerSerLeuProSerAsnAspTyrTrpGlyAspThrMetArgLeu-----TyrLysGly 66  
:: ::  
124 TCACAGCTTTCACAGAAAGCTTGGAGACCGAAAGAGCCCTGATCGAGTATGTTGGT 183  
DB  
67 PheTrpGlnMetGlyTyrLeuValProGlyIleMetalPheGluAspAsnPheLysAla 86  
:: ::  
184 TACTGGTGCTACCGTCTCTCCTCGAAGTGTATTCAACGCGCAAGAGTTCTTTCAAGCA 243  
QY  
87 ArgGluThrAspIleLeuThrThrLeuProLysAlaGlyThrThrTrpThrLysAla 106  
:: ::  
244 CGACCCAGTGACTTCTCTGTGTGATGCTACCAAGACGAGCACCATCTGGCTCAAAAGCC 303  
DB  
107 LeuThrPheAlaIleLeuThrArgAspValAsnHisProSerSerProThrHisProLeu 126  
:: ::  
304 CTGACHTTCGCATCGCAATCGTCCCGCTTCGATGATCTCTCC-----AACCTCTC 357  
DB  
127 LeuPhePheAsnProHisSerCysValGlnAsnLeuGluTyrLeuTyrMetGlyArgGlu 146  
:: ::  
358 CTGAAACGTAACCCCTCACGAGTTTGTCTCTACATTGAGATAGATTTC----- 405  
QY  
147 AsnThrMetProAspLeuAspMetLeuAsnGluSerProArg-----LeuPheAlaGlyHis 165  
:: ::  
406 CCTTCTTCCCTCGAAGTGTATGTTCTCAAGACAAAGGGAACATCTGTTTTGCATCAT 465  
QY  
166 IleProTyrSerLeuLeuProAlaSerValLeuLysSerGlyThrLysIleIleAsnIle 185  
:: ::  
466 ATCCATACAGATTATTACTGATTGGTTGTGAAATCCGGTTGAAGATGGTTACATA 525  
DB  
186 SerArgAsnArgLysSerThrPheValSerPheThrLysPheGlyAsnLeuIleAsnPro 205  
:: ::  
526 TGGAGAGAACCAAGACACTTTCATCTCCATGTGGAGCTTTC-----CTTCACAAG 576  
DB  
206 AspLys-----LeuLeuAspLeuGluLysSerValAspIlePheAlaSer 220  
:: ::  
577 GAAAGACAGAGCTTGGACCTCTCAGCAATCTTGAGGAGTCTTTTGATATGTTCTGTGCT 636  
DB  
221 GlyIleSerPheCysGlyProGluTrpAsnPheGlnAlaGluPheThrAsnAlaAlaSer 240  
:: ::  
637 GGTCTGTCTGGGTATGTTCTTATCTTAATCATATCCTCTGGCGTATTGGAAAGCATACCA 656  
QY  
241 ThrAsn---SerAsnLeuLeuLeuLeuSerTyrGluGluMetLeuGluLysProValGlu 259  
:: ::  
697 GAGAAATCCAGATAGGATCTTGTCTCAAGTACGAGACGATGAGAGCTGTATCTTTTACCG 756  
DB  
260 AsnValLysLysLeuAlaGluPheMetGlyCysGlyPheThrAspAspGluGluLysGln 279  
:: ::  
757 TAGTGAAGAGTCTGGCTGAGTTTATGGGTCTATGGATTTCACAGCCGAGGAGAGAGAAA 816  
DB  
280 GlyIleValAspGluIleValLysLeuCysSerPheAspAsnLeuLysAsnGlnGlnVal 299  
:: ::  
817 GGTGTGTTTCAGAAATGGTGAATCTTTTCACCTTCGAGAGCTTGAGAAATCTTTGAAGCT 876  
DB

300	QY	AenLysAenGlySerSerTyrAsnSerLys-----IleAspAsnLysHisPhePhe	316
		:::	:::
877	Db	AACAAAGGGGAGAAAGACAGAGAGATCGCTCGTGTTTACGCGAATAGCGGTATTTTC	936
		:::	:::
317	QY	ArgLysGlyGluValAlaGAspTTPAlaAsnTyrLeuThrSerGluMetIleLysLysLeu	336
		:::	:::
937	Db	AGCAAGAAAGAGTGGGAGATTGGTCGACTATCTGACTCCGGAGATGGTGTCTCGTATA	996
		:::	:::
337	QY	GluThrAlaGlyLysIleAsnGluSerGluLys	347
		:::	:::
997	Db	GAT-----GGTTAATGGAAGAGAAATTTAAG	1023
		:::	:::
RESULT 8			
AAD54410		ID	AAD54410 standard; cDNA; 1400 BP.
XX		AC	AAD54410;
XX		AC	
XX		DT	17-JUN-2003 (first entry)
XX		DE	Lolium perenne defensin e (LpDeFe) cDNA.
XX		XX	
KW		KW	Ryegrass; thionin; TH; thaumatin-like protein; TL; defensin; DEF; ER;
KW		KW	elicitor-responsive protein; disease resistance; plant defence response;
KW		KW	protein storage; pest resistance; genetic marker; gene therapy; antipept;
KW		KW	agricultural; LpDeFe; gene; ss.

XX Lolium perenne.

XX	Key	Location/Qualifiers
PH	FT	153..833
PH	FT	/*tag= a
PH	FT	/product= "RveCrass DEFE (pDEFE) protein"

AA PN WO200288359-A1

XX  
07-NOV-2002XX  
PF 01-MAY-2002: 2002WO-AIJ000539.

02-MAY-2001: 2001AII-00004735

XX  
BA (ACPI-) AGRIC VICTORIA SERVICES PTY LTD.PA (AGRE-) AGRESEARCH LTD.  
yy

PI Spangenberg G, Sawbridge TI, Ong EK, Emmerling M;  
yy

DR WPI; 2003-201227/19.  
DD D. DCDR. XAE3E047

**XX**

PT elicitor-responsive or defensin polypeptide, for modifying disease and/or  
PT plant resistance, plant defence response and/or protein storage in a  
PT plant.

XX  
33  
Class 6. Page 62. 1950. English

xx The invention relates to novel ryegrass (*Lolium*) or fescue (*Festuca*)  
cc nucleic acids encoding thionin (TH), thaumatin-like (TL), elicitor-  
cc responsive (ER) or defensin (DEF) proteins and polynucleotides encoding  
cc such proteins. Sequences of the invention are useful to modify disease  
cc and/or pest resistance, plant defence response and/or protein storage in  
cc plants. The nucleotide sequence information and/or single nucleotide  
cc polymorphisms of the nucleic acid are useful as genetic markers. They are  
cc also useful in gene therapy. The present sequence is ryegrass DEFc  
cc (LoDEFc) cDNA

XX  
SQ  
Sequence 1400 BP: 319 A: 409 C: 380 G: 292 T: 0 U: 0 Other:

Alignment Scores:  
Pred. No.:  
Score:

Db	1004	ATGCCCATGGAGAAGTCCTCTACTCAGGAATGGAGCGGTGGCGGATGGGCCAACCCAC	1063
Qy	328	LeuThrSerGluMetIleLysLysLeuGluThrAlaGlyLysIleAsnGluSerGluLys	347
Db	1064	CTGACCCAGAGATGGCCACAGTTGGAT-----GGCATCGTCAGGAGAGCTCAAA	1117
Qy	348	-----	HisLeu 349
Db	1118	GGTTCTGCCTCGCCTTTTGAGATTTCTTCTCCGATCGATCTGTGAGATTTCTATAT	1177
Qy	350	LeuSerAsnLysAsnLeuThrAsnPheAsnLeuAsnAsnTyrCysGluLysSerAsnIle	369
Db	1178	CTCGTGGCCCAATATCTCATCTCAATTAATACTATT--GGATCATGTTATGTTTCTACTATC	1234
Qy	370	SerLeuCysLeuLeuSerTyrSerPheVal-----IleAsnAsnPheIleLeuLeuLys	387
Db	1235	-----GTTATGTTCTTATGGAATAATGTGAAGACCAAGTCGCTGGTCTATTGCA	1285
Qy	388	LysLysLysLys 391	
Db	1286	AAAAAAAAAAAA 1297	
RESULT 9			
AAAC37521			
ID	AAAC37521	standard; DNA; 1205 BP.	
XX	AAAC37521;		
AC	AAAC37521;		
XX	17-OCT-2000	(first entry)	
XX	Arabidopsis thaliana	DNA fragment SEQ ID NO: 17688.	
XX	Hybridisation assay; genetic mapping; gene expression control;		
XX	protein identification; signal transduction pathway; metabolic pathway;		
XX	promoter; termination sequence; ss.		
XX	Arabidopsis thaliana.		
XX	EP1033405-A2.		
PD	06-SEP-2000.		
XX	25-FEB-2000; 2000EP-00301439.		
XX	25-FEB-1999;	99US-0121825P.	
PR	05-MAR-1999;	99US-0123180P.	
PR	09-MAR-1999;	99US-0123548P.	
PR	23-MAR-1999;	99US-0125788P.	
PR	25-MAR-1999;	99US-0126264P.	
PR	29-MAR-1999;	99US-0126785P.	
PR	01-APR-1999;	99US-0127462P.	
PR	06-APR-1999;	99US-0128234P.	
PR	08-APR-1999;	99US-0128714P.	
PR	16-APR-1999;	99US-0129845P.	
PR	19-APR-1999;	99US-0130077P.	
PR	21-APR-1999;	99US-0130449P.	
PR	23-APR-1999;	99US-0130510P.	
PR	23-APR-1999;	99US-0130891P.	
PR	28-APR-1999;	99US-0131449P.	
PR	30-APR-1999;	99US-0132048P.	
PR	30-APR-1999;	99US-0132407P.	
PR	04-MAY-1999;	99US-0132484P.	
PR	05-MAY-1999;	99US-0132485P.	
PR	06-MAY-1999;	99US-0132486P.	
PR	06-MAY-1999;	99US-0132487P.	
PR	07-MAY-1999;	99US-0132863P.	
PR	11-MAY-1999;	99US-0134256P.	
PR	14-MAY-1999;	99US-0134218P.	
PR	14-MAY-1999;	99US-0134219P.	
PR	14-MAY-1999;	99US-0134221P.	
PR	14-MAY-1999;	99US-0134370P.	
PR	18-MAY-1999;	99US-0134768P.	
PR	19-MAY-1999;	99US-0134941P.	

Percent Similarity:	50.47%	Conservative:	65
Best Local Similarity:	35.21%	Mismatches:	154
Query Match:	27.10%	Indels:	58
DB:	7	Gaps:	12
US-09-854-122-16 (1-391) x AAD54410 (1-1400)			
Qy	1	ThrArgGlyIleThrGlyIleAlaValAlaCysLeuProLeuIleMetAlaGlyIleLeu	20
Db	93	ACGCRAGTAGTCGGGAGCTCAGATTAGACTAGCACACCACCTCGCCGGA-----	146
Qy	21	AlaLeuGluLysCysPheGlySerLysAsnGluGlnGluLysGluAspSerLysMet	40
Db	147	-----TTGCGCATGGATCCAACTCAAAACCAAGACGCGTCGCAATGCC	191
Qy	41	TyrLysArgTyrArgGluIleValSerSerLeuProSerAsnAspTyrTyrGlyAspThr	60
Db	192	GCCGAGAAAGAGAGCGCTTGTGTCCAGCTCCCATGAGGAGAGATCTCGAAACCG	251
Qy	61	MetArgLeuTyrLysGlyPheTyrGlnMetGlyTyrLeuValProGlyIleMetAlaPhe	80
Db	252	TACGTCTCTACCGAGGCTTCTGGGTATCTCCGAGGTGGCCAAAGAGCGTCATGCTTG	311
Qy	81	GluAspAsnPheLysAlaArgGluThrAspIleIleLeuThrThrLeuProLysAlaGly	100
Db	312	CAAGACGAGTCAAGCGCGCGCGACGACATCATCTCCGCGGTCCGCCCAAGTCCGCG	371
Qy	101	ThrThrThrThrLysAlaLeuThrPheAlaIleLeu-Thr---ArgAspValAsnHisPr	119
Db	372	ACCACCTGGTCAAGCGCTTCCCTTACCCTCGTCAACCGCTCGCACGCGCGTGGCC	431
Qy	119	oSerSerProThr---HisProLeuLeuPhePheAsnProHisSerCysValGlnAsnLeu	138
Db	432	GGAGACGCGGACGCGCACCGCTGCTTACCACACGCCCCCATGACCTCGTCCGCTTATC	491
Qy	139	GluTyrLeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMet-----	154
Db	492	GAG-----AAACCTGACCGGAGCTCTACCTGTGCGCC	524
Qy	155	-----LeuAsnGlu-----SerProArgLeuPheAlaGlyHisIleProTyrSerLeuLeu	171
Db	525	GAACCTGGAGGAGCTCGCGTCCCAAGGCTCCTGGCGGACCCACATGCGCTTCCGCTGTG	584
Qy	172	ProAlaSerValLeuLysSerGlyThrLysIleIleAsnIleSerArgAsnArgLysSer	191
Db	585	CCGGAGAGATCTCCGCGCTCGGCTCGCGTGGGTATACATCAGCCGGAGGCCAAGGAT	644
Qy	192	ThrPheValSerPheThrLysPheGlyAsnLeuIleAsnProAspLysLeuLeuAspLeu	211
Db	645	GTGCTCGTCTCGCTGGCACTACTGAACTCGGTAAAGCAAGACCA-CTTCATCAGCTG	703
Qy	212	GluLysSerValAspIlePheAlaSerGlyIleSerPheCysGlyProGluTyr---Asn	230
Db	704	GAGAAAGCCTTCAGTGTTCAGCAGGAGGGGTCTATTTTCGTCAGTGTGGATCAC	763
Qy	231	PheGlnAlaGluPheThrAsnAlaAlaSerThrAsnSerAsnLeuLeuLeuSerTyr	250
Db	764	TATCTTTGGTACTGGAACAGACGATTCGAGACGCCAGATCGGGTCTCTCTCTCAAGTAC	823
Qy	251	GluGluMetLeuLysProValGluAsnValLysLysLeuAlaGluPheMetGlyCys	270
Db	824	GAGGATGATGCGCCACCGCGTCAAGCAGCTCAGACCTCGCCGAGTTCCTTGGATC	883
Qy	271	GlyPheThrAspAspGluGluLysGlnGlyIleValAspLysLeuValLysLeuCysSer	290
Db	884	CCGTTACCGTCCAGCGGTGGAGCGCGGAGCTGTGGAGCAGGTCTGGACCTGTGTAGC	943
Qy	291	PheAspAsnLeuLysAsnGlnGlnValAsnLysAsnGlySerSerTyrAsnSer-----	308
Db	944	TTTCAGAGCTCAAGAACCTGCTCACTCTCGGGGACATCCAACTCGCTGGTGG	1003
Qy	309	---LysIleAspAsnLysHisPhePheArgLysGlyValArgAspThrAlaSerTyr	327

us-09-854-122-16.rng

Mon Mar 22 09:47:42 2004

PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138947P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 24-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 28-JUN-1999; 99US-0140891P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142330P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143354P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144844P.  
PR 21-JUL-1999; 99US-0144844P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145132P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147432P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149358P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149240P.  
PR 25-AUG-1999; 99US-0150586P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 30-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157855P.  
PR 07-OCT-1999; 99US-0158023P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159323P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 23-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.

PR 26-OCT-1999; 99US-0161361P.  
 PR 28-OCT-1999; 99US-0161920P.  
 PR 28-OCT-1999; 99US-0161932P.  
 PR 28-OCT-1999; 99US-0161933P.  
 PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:  
 Pred. No.: 1,03e-44  
 Score: 557.50  
 Percent Similarity: 56.23%  
 Best Local Similarity: 37.67%  
 Query Match: 27.08%  
 DB: 3  
 Gaps: 9

US-09-854-122-16 (1-391) x AAC37521 (1-1205)

QY 29 LysAsnGluGlnGluLysGluLysSerLysMetTyrLysArgTyrArgGluLeuVal 48  
 DB 64 GAAGAAGAGAAACCAAGTGAAGATTCAAA-----AGTTTGATC 105

QY 49 SerSerLeuProSerAsn---AspTyrTrpGlyAspThrMetArgLeuTyrLysGlyPhe 67  
 DB 106 TCTTCATCTCTTCAGACATAGATTGCTCTGGAGCAAGTTGTACAACTATCAAGATGAT 165

QY 68 TrpGlnMetGlyTyrLeuValProGlyIleMetAlaPheGluAspAsnPhelLysAlaArg 87  
 DB 166 TGGTACGATTAAGACATCTCCAGCAATCTCCAAATTCACAAATCTTTCAGCCACAA 225

QY 88 GluThrAspIleLeuThrThrLeuProLysAlaGlyThrThrTrpThrLysAlaLeu 107  
 DB 226 GAAACCGATATATGTTGCTCTTCCCAATCAATCAGTACAGTCTGCTCAAGCATC 285

QY 108 ThrPheAlaLeuThrArgAspValAlaHisProSerSerProThrHisProLeuLeu 127  
 DB 286 ACATTCGCATCTCGCTCAAGATCAAAACAT-----ACTTCAGAAATCATCTCTGCTA 339

QY 128 PhePheAsnProHisSerCysValGlnAsnLeuGluTyr---LeuTyrMetGlyArgGlu 146  
 DB 340 ACTCATATCTCATGAGCTAGTGCCTACCTCGACTCGATCTTATCTC-----AAA 393

QY 147 AsnThrMetProAspLeuAspMetLeu---AsnGluSerProArgLeuPheAlaGlyHis 165  
 DB 394 AGTCGAAACCGGATATGTCAGATGATCCATCATCTCCGAGATGTTCTCAACCCAC 453

QY 166 IleProTyrSerLeuLeuProAlaSerValLeuLysSerGlyThrLysIleLeuAsnIle 185  
 DB 454 ATGCTCTTCGATCGCTTAAAGTACCAATCAAGGACACTCTTCGACAGATAGTATGTG 513

QY 186 SerArgAsnArgLysSerThrPheValSerPheTrpLysPheGlyAsnLeuLeuAsnPro 205  
 DB 514 TGCAGGAACGTAAAGACGCTGGTGGTATCACTTGGTGTTCGAAACTCCATTAGTGA 573

QY 206 AspLysLeuLeuAspLeuGluLysSerValAspIlePheAlaSerGlyLysSerPheCys 225  
 DB 574 GAAACAAATTAAGTCTCGAGGCTTTGTTTCGAGTCTTTATGTCGAGGATTAACCTATGC 633

QY 226 GlyProGluTrp---AsnPheGlnAlaGluPheThrAsnAlaSerThrAsnSerAsn 244  
 DB 634 GGTCCTTGTGGGAAATATGTTAGGCTATTGAGAGGAGGCTTGGAGATCTTAAGCAT 693

QY 245 LeuLeuLeuLeuSerTyrGluGluMetLeuGluLysProValGluAsnValLysLysLeu 264  
 DB 694 GTGCTTCTTGTAGGTACGAGGATTTGAAGACGAGCCCTGCTGCAAAATCAAGACTT 753

QY 265 AlaGluPheMetGlyCysGlyPheThrAspAspGluGluLysGlnGlyIleValAspGlu 284  
 DB 754 GCAGAGTCTTGAATTTCCATTCACAAAGGAGGAGAGATAGTGGAGGTGTAGACAA 813

QY 285 IleValLysLeuCysSerPheAspAsnLeuLysAsnGlnGlnValAsnLysAsnGlySer 304  
 DB 814 ATCTTGAACCTTGTCTCTAAGAAACCTTAGCGGTTTCGAGATCAACAAACACAGGA-- 870

QY 305 SerTyrAsnSerLysIleAspAsnLysHisPheArgLysGlyGluValArgAspTrp 324

DB 871 AGCTTGTCCGAAGGAGTAAGTTTCAAGAGATTTTTCGTAAGGGAAGTGGTGGTATGG 930  
 QY 325 AlaAsnTyrLeuThrSerGluMetIleLysLysLeuGluThrAlaGlyLysIleAsnGlu 344  
 DB 931 AAGAGTTATATGCTCTCTGAAATGGAACAAATCGAC-----ATGATTTGT 978

QY 345 SerGluLysHisLeuLeuSerAsnLysAsnLeuThrAsnPheAsnLeuAsnTyrCys 364  
 DB 979 GAGGAGAAACTTCAAGGCTCTGGTTTGATTTAGATTTAGAGTTCATCTCTATGTATGTG 1038

QY 365 GluLysSerAsnIle-SerLeuCysLeuLeuSerTyrSerPheValIleAsnAsnPheIle 384  
 DB 1039 AACAGGTTTAAATCTCAACCTAAATATGCTGGTTGTTCTTTTCTTGTATCAATGTAAT 1098

QY 384 e 384  
 DB 1099 A 1099

RESULT 10  
 AAF29177  
 ID AAF29177 standard; DNA; 1077 BP.  
 XX AAF29177;  
 AC AAF29177;  
 XX 09-APR-2001 (first entry)  
 XX Hydroxyjasmonic acid sulfoltransferase AtST2a gene.  
 XX Hydroxyjasmonic acid sulfoltransferase; AtST2a; flowering time; cabbage;  
 XX jasmonate; genetically modified plant; lettuce; sugar cane; carrot;  
 XX increase vegetative growth; biomass increase; ds.  
 XX Arabidopsis thaliana.  
 XX WO200102589-A2.  
 XX 11-JAN-2001.  
 XX 06-JUL-2000; 2000WO-CA000801.  
 XX 06-JUL-1999; 99CA-02274873.  
 XX (VARI/) VARIN L.  
 XX (GIDD/) GIDDA S.  
 XX Varin L, Gidda S;  
 XX WPI; 2001-159272/16.  
 XX P-PSDB; AAB49722.

Methods for modulating flowering in plants, particularly useful for plants used in the food-processing industry, involves modifying the endogenous level of compounds of the jasmonate family.

Claim 38; Fig 7; 50pp; English.

This invention relates to a method for modulating flowering in a plant. The method comprises modifying the endogenous level of at least one compound of the jasmonate family in a plant. The methods are used to produce plants which are genetically modified to flower early or tardily when compared to a corresponding plant that is not genetically modified, where the modified plant has an increased (flower early) or lowered (flower tardily) level of jasmonic acid or a compound of the jasmonate family. The method is useful for modulating flowering, particularly for plants that are used in the food-processing industry and plants with horticultural value. The method is particularly useful for e.g. delaying flowering time in crops like lettuce, cabbage, sugar cane or carrots, which results in increased vegetative growth and biomass. The present sequence represents the Arabidopsis thaliana AtST2a gene, which encodes an hydroxyjasmonic acid sulfoltransferase protein, which can be used in the method of the invention





PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139454P.  
PR 17-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 22-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 03-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.

PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 18-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151503P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154179P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161922P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:

Pred. No.: 1,74e-44

Score: 555.50

Percent Similarity: 55.98%

Best Local Similarity: 35.86%

Length:

1270

Matches:

123

Conservative:

69

Mismatches:

140

Query Match:	26.98%	Indels:	11
DB:	3	Gaps:	8
US-09-854-122-16 (1-391) x AAC48839 (1-1270)			
QY	1	ThrArgGlyIleThrGlyIleAlaValAlaCysLeuProLeuIleMetAlaGlyIleLeu	20
DB	70	ACCTCAAGCATGAAGAGCATTCCTCAATGGCGATCCCAAGTTTCTCCATGTGTCAACAAGCTC	129
QY	21	AlaLeuGluLysCysPheGlySerLysAsnGluGlnGluLysGluGluAspSerLysMet	40
DB	130	GAGTCTCTTAAGAGGCAAACTCCGACGTCGCCGAAGCCGAAGAGATGAAGGGCTA	189
QY	41	TyrLysArgTyrArgGluIleValSerSerLeuProSerAsnAspTyrTyrGlyAspThr	60
DB	190	AGTGGAGTTCGAAGAGATGTGGATTCTCTTCTTAAGGAGAGAGATGGAGAACTCGT	249
QY	61	MetArgLeuTyrLysGlyPheTyrGlnMetGlyTyrLeuValProGlyIleMetAla	79
DB	250	TACCTTTTACCTATTCGAAGGGTTTGTGTCGCAAGCCCAAGAGATTCAGCCATCATGTCT	309
QY	80	PheGluAspAsnPhelysAlaArgGluThrAspIleIleLeuThrThrLeuProLysAla	99
DB	310	TTCCAAAACATTTCCAAATCCCTCGAAACAGCGTCTCTCCGACCATCAATATCC	369
QY	100	GlyThrThrTyrLysAlaLeuThrPheAlaIleLeuThrArgAspValAsnHisPro	119
DB	370	GGTACAACTGGTAAAGCTTTTAACTTTCACCTTCCACCTCATGACCTTGCTTTC	429
QY	120	-----SerSerProThrHisProLeuLeuPheAsnProHisSerCysValGlnAsn	137
DB	430	GTTGCTCGAGTACCAACCACTCTTTTCACTTCCACCTCATGACCTTGCTTTC	489
QY	138	LeuGluTyr---LeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsn	156
DB	490	TTGAGTACAAAGCTTACGCCAACGAGAT-----GTTCCCGATCTCCGGGTCTA---	540
QY	157	GluSerProArgLeuPheAlaGlyHisIleProTyrSerLeuLeuProAlaSerValLeu	176
DB	541	GCAGTCCAGAGAGCTTGGCAACCACTTACCGTTCCGTTCCCTAAGGAACAGTACGAG	600
QY	177	LysSerGlyThrLysIleIleAsnIleSerArgAsnArgLysSerThrPheValSerPhe	196
DB	601	AAACCCGGTGTGAAGTCTGTACTTGTGCGGACCCGTTTGACACATTCATCTCTTCG	660
QY	197	TyrLysPheGlyAsnLeuIleAsnProAspLysLeu-----LeuAspLeuGluLysSer	214
DB	661	TGGCATTACCAACCAACATCAATCCGATCGATCGAGCCGCTGTGTAGACCAAGCT	720
QY	215	ValAspIlePheAlaSerGlyIleSerPheCysGlyProGluTyrAsnPheGlnAlaGlu	234
DB	721	TTTGATCTGTATTCGCGGAGTGTGCGGTTTGGCCGTTTGGGAACACATGTTGGGA	780
QY	235	Phe---ThrAsnAlaAlaSerThrAsnSerAsnLeuLeuLeuSerTyrGluGluMet	253
DB	781	TACTGGAGAGAGCTTGAAGAGACGACGAGAAAGTCTCTTTTAAAGTACGAGATCTC	840
QY	254	LeuGluLysProValGluAsnValLysLysLeuAlaGluPheMetGlyCysGlyPheThr	273
DB	841	AAAGACGACATCGAGACCACTTGAAGAGCTTTCGCACTTCTTAGAGCTTCTTCACC	900
QY	274	AspAspGluLysGlnGlyIleValAspGluIleValLysLeuCysSerPheAspAsn	293
DB	901	GAAGAAGAGAAAGAGGAGTGTGAAGCTATCCGCGAGCTGTAGCTTCGAGAT	960
QY	294	LeuLysAsnGlnGlnValAsnLysAsnGlySerSerTyrAsnSerLysIleAspAsnLys	313
DB	961	CTGAGAGTGTGAGGTGAACAGTCAACAGTCAAGTCAAGTCAAGTCAAGTCAAGT	1017
QY	314	HisPhePheArgLysGlyGluValArgAspTyrPalAsnTyrLeuThrSerGluMetIle	333
DB	1018	TTCTTGTTCGGAAGAGAGAGTGAAGTGTGGTAACTATTTGTCACCTTCACAGT	1077

QY	334	LysLysLeu	336
DB	1078	GAAGATTG	1086
RESULT 12			
AACT7848			
ID	AAC47848	standard; DNA; 1273 BP.	
XX	AC	AAC47848;	
XX	18-OCT-2000	(first entry)	
XX	Arabidopsis thaliana	DNA fragment SEQ ID NO: 55342.	
XX	Hybridisation assay; genetic mapping; gene expression control;		
KW	protein identification; signal transduction pathway; metabolic pathway;		
KW	promoter; termination sequence; ss.		
XX	Arabidopsis thaliana.		
XX	EPI033405-A2.		
XX	06-SEP-2000.		
XX	25-FEB-2000; 2000EP-00301439.		
PR	25-FEB-1999;	99US-0121825P.	
PR	05-MAR-1999;	99US-0123180P.	
PR	09-MAR-1999;	99US-0123548P.	
PR	23-MAR-1999;	99US-0125788P.	
PR	25-MAR-1999;	99US-0126264P.	
PR	29-MAR-1999;	99US-0126785P.	
PR	01-APR-1999;	99US-0127462P.	
PR	06-APR-1999;	99US-0128234P.	
PR	08-APR-1999;	99US-0128714P.	
PR	16-APR-1999;	99US-0129845P.	
PR	19-APR-1999;	99US-0130077P.	
PR	21-APR-1999;	99US-0130449P.	
PR	23-APR-1999;	99US-0130510P.	
PR	28-APR-1999;	99US-0130891P.	
PR	30-APR-1999;	99US-0131449P.	
PR	30-APR-1999;	99US-0132048P.	
PR	04-MAY-1999;	99US-0132407P.	
PR	05-MAY-1999;	99US-0132484P.	
PR	06-MAY-1999;	99US-0132485P.	
PR	06-MAY-1999;	99US-0132486P.	
PR	07-MAY-1999;	99US-0132487P.	
PR	11-MAY-1999;	99US-0132863P.	
PR	14-MAY-1999;	99US-0134256P.	
PR	14-MAY-1999;	99US-0134218P.	
PR	14-MAY-1999;	99US-0134219P.	
PR	14-MAY-1999;	99US-0134221P.	
PR	14-MAY-1999;	99US-0134370P.	
PR	18-MAY-1999;	99US-0134768P.	
PR	19-MAY-1999;	99US-0134941P.	
PR	20-MAY-1999;	99US-0135124P.	
PR	21-MAY-1999;	99US-0135353P.	
PR	24-MAY-1999;	99US-0135629P.	
PR	25-MAY-1999;	99US-0136021P.	
PR	27-MAY-1999;	99US-0136392P.	
PR	28-MAY-1999;	99US-0136782P.	
PR	01-JUN-1999;	99US-0137222P.	
PR	03-JUN-1999;	99US-0137528P.	
PR	10-JUN-1999;	99US-0137502P.	
PR	07-JUN-1999;	99US-0137724P.	
PR	08-JUN-1999;	99US-0138094P.	
PR	10-JUN-1999;	99US-0138540P.	
PR	10-JUN-1999;	99US-0138847P.	
PR	14-JUN-1999;	99US-0139119P.	
PR	16-JUN-1999;	99US-0139452P.	
PR	16-JUN-1999;	99US-0139453P.	
PR	17-JUN-1999;	99US-0139492P.	
PR	18-JUN-1999;	99US-0139454P.	



Db 72 ACCTCAACGTCAGAGAGCATTCCTCAATGCGGATGCCATCCCAAGTTTCTCCATGTCACAGGTC 131  
Qy 21 AlaLeuGluLysCysPheGlySerLysAsnGluGlnGluLysGluAspSerLysMet 40  
Db 132 GAGCTCTCTAAAGAGGCAAACTCGCAGCTCCGAAAGCGGAGAGATGAAGGGCTA 191  
Qy 41 TyrLysArgTyrArgGluLeuValSerSerLysLeuProSerAsnAspTyrTyrGlyAspThr 60  
Db 192 ACCTCGAGATTCCTCAAGAGATGTGGATCTCTCTTAAGGAGAGAGATGAGAACTCGT 251  
Qy 61 ---MetArgLeuTyrLysGlyPheTyrGlnMetGlyTyrLeuValProGlyLeuMetAla 79  
Db 252 TACCTTTACCTATTCCAAAGGGTTTGGTGCCAAAGCAAGAGATTCAAGCCCATCATGTCT 311  
Qy 80 PheGluAspAsnPhelysAlaArgGluThrAspIleLeuThrThrLeuProLysAla 99  
Db 312 TTCCAAAACATTCATCCCTCGAAAGACGCTGCTCGCCACCATCACTAAATCC 371  
Qy 100 GlyThrThrTyrLysAlaLeuThrPheAlaIleLeuThrArgAspValAsnHisPro 119  
Db 372 GGTACAACTGGCTAAAGCTTTAACTTTCACCATCTTAACCGTCACCGGTTTGATCCG 431  
Qy 120 -----SerSerProThrHisProLeuLeuPhePheAsnProHisSerCysValGlnAsn 137  
Db 432 GTTGCTCGAGTACCAACACCTCTTTTCACTTCCACCTCATGACCTTGACTTTC 491  
Qy 138 LeuGluTyr---LeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsn 156  
Db 492 TTCGAGTACAGCTTTACGCCACCGAGAT-----GTTCCTCGATCTCTCGGGTCTA--- 542  
Qy 157 GluSerProArgLeuPheAlaGlyHisIleProTyrSerLeuLeuProAlaSerValLeu 176  
Db 543 GCCAGTCCAAAGACGTTCCCAACCTTACCGTTCCGTTTAAAGAAACGATCGAG 602  
Qy 177 LysSerGlyThrLysIleIleAsnIleSerArgAsnArgLysSerThrPheValSerPhe 196  
Db 603 AAACCGGTGTGAAGTCTGCTGTTGCGGAAACCGGTTTGACACATTCATCTCTTCG 662  
Qy 197 TrpLysPheGlyAsnLeuLeuAsnProAspLysLeu-----LeuAspLeuGlyLysSer 214  
Db 663 TGGCATTTACCAACCAACATCAATCCGAGTCAGTCCAGCCAGCTCTTGCTAGACCAAGCT 722  
Qy 215 ValAspIlePheAlaSerGlyIleSerPheCysGlyProGluTrpAsnPheGlnAlaGlu 234  
Db 723 TTGTATCTGTATTCGCGGAGTGATCGGGTTTGGCCCGTTTGGGAAACACATGTTGGGA 782  
Qy 235 Phe---ThrAsnAlaAlaSerThrAsnSerAsnLeuLeuLeuLeuSerTyrGluGluMet 253  
Db 783 TACTGGAGAGAGAGCTTGAAGAGACCAAGAGAAAGTCTCTTTTAAAGGTACGAGGATCTC 842  
Qy 254 LeuGluLysProValGluAsnValLysLeuAlaGluPheMetGlyCysGlyPheThr 273  
Db 843 AAAGACGACATCGAGACCACTTGAAGAGCTTGCAACTTCTTAGAGCTTCTTTCACC 902  
Qy 274 AspAspGluGluLysGlnGlyIleValAspGluIleValLysLeuCysSerPheAspAsn 293  
Db 903 GAGAAAGAGGACCAAGAGGAGTTGTGAAGCTATCCGCGAGCTGTGTAGCTTCGAGAAT 952  
Qy 294 LeuLysAsnGlnGlnValAsnLysAsnGlySerSerTyrAsnSerLysIleAspAsnLys 313  
Db 963 CTGAAGAAGTTGGAGGTGAACAGTCAACAAAGTCGATCAAGAAC---TTTGAGATCGA 1019  
Qy 314 HisPhePheArgLysGlyGluValArgAspTrpAlaAsnTyrLeuThrSerGluMetIle 333  
Db 1020 TTCTTGTTTCGAAGGAGAGTGAAGTGTGGTGAATTTGTTGTTCACTTGTACCTTCAAGTG 1079  
Qy 334 LysLysLeu 336  
Db 1080 GAAAGATTG 1088

RESULT 13  
AAC37672  
ID AAC37672 standard; DNA; 1154 BP.

XX AAC37672;  
XX 17-OCT-2000 (first entry)  
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 18246.  
XX Hybridisation assay; genetic mapping; gene expression control;  
XX protein identification; signal transduction pathway; metabolic pathway;  
XX promoter; termination sequence; ss.  
XX Arabidopsis thaliana.  
XX EP1033405-A2.  
XX 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-00301439.  
XX 25-FEB-1999; 99US-0121825P.  
XX 05-MAR-1999; 99US-0123180P.  
XX 09-MAR-1999; 99US-0123548P.  
XX 23-MAR-1999; 99US-0125788P.  
XX 25-MAR-1999; 99US-0126264P.  
XX 29-MAR-1999; 99US-0126785P.  
XX 01-APR-1999; 99US-0127462P.  
XX 06-APR-1999; 99US-0128234P.  
XX 08-APR-1999; 99US-0128714P.  
XX 16-APR-1999; 99US-0129845P.  
XX 19-APR-1999; 99US-0130077P.  
XX 21-APR-1999; 99US-0130449P.  
XX 23-APR-1999; 99US-0130510P.  
XX 23-APR-1999; 99US-0130891P.  
XX 28-APR-1999; 99US-0131449P.  
XX 30-APR-1999; 99US-0132048P.  
XX 30-APR-1999; 99US-0132407P.  
XX 04-MAY-1999; 99US-0132484P.  
XX 05-MAY-1999; 99US-0132485P.  
XX 06-MAY-1999; 99US-0132486P.  
XX 06-MAY-1999; 99US-0132487P.  
XX 07-MAY-1999; 99US-0132863P.  
XX 11-MAY-1999; 99US-0134256P.  
XX 14-MAY-1999; 99US-0134218P.  
XX 14-MAY-1999; 99US-0134219P.  
XX 14-MAY-1999; 99US-0134221P.  
XX 14-MAY-1999; 99US-0134370P.  
XX 18-MAY-1999; 99US-0134768P.  
XX 19-MAY-1999; 99US-0134941P.  
XX 20-MAY-1999; 99US-0135124P.  
XX 21-MAY-1999; 99US-0135353P.  
XX 24-MAY-1999; 99US-0135629P.  
XX 25-MAY-1999; 99US-0136021P.  
XX 27-MAY-1999; 99US-0136392P.  
XX 28-MAY-1999; 99US-0136782P.  
XX 01-JUN-1999; 99US-0137222P.  
XX 03-JUN-1999; 99US-0137528P.  
XX 04-JUN-1999; 99US-0137502P.  
XX 07-JUN-1999; 99US-0137724P.  
XX 08-JUN-1999; 99US-0138094P.  
XX 10-JUN-1999; 99US-0138540P.  
XX 10-JUN-1999; 99US-0138847P.  
XX 14-JUN-1999; 99US-0139119P.  
XX 16-JUN-1999; 99US-0139452P.  
XX 16-JUN-1999; 99US-0139453P.  
XX 17-JUN-1999; 99US-0139492P.  
XX 18-JUN-1999; 99US-0139454P.  
XX 18-JUN-1999; 99US-0139455P.  
XX 18-JUN-1999; 99US-0139456P.  
XX 18-JUN-1999; 99US-0139457P.  
XX 18-JUN-1999; 99US-0139458P.  
XX 18-JUN-1999; 99US-0139459P.  
XX 18-JUN-1999; 99US-0139460P.  
XX 18-JUN-1999; 99US-0139461P.

PR	25-AUG-1999;	99US-0150568P.
PR	26-AUG-1999;	99US-0150884P.
PR	27-AUG-1999;	99US-0151065P.
PR	27-AUG-1999;	99US-0151066P.
PR	27-AUG-1999;	99US-0151080P.
PR	30-AUG-1999;	99US-0151303P.
PR	31-AUG-1999;	99US-0151438P.
PR	01-SEP-1999;	99US-0151930P.
PR	07-SEP-1999;	99US-015263P.
PR	10-SEP-1999;	99US-0153070P.
PR	13-SEP-1999;	99US-0153758P.
PR	15-SEP-1999;	99US-0154018P.
PR	16-SEP-1999;	99US-0154039P.
PR	20-SEP-1999;	99US-0154779P.
PR	22-SEP-1999;	99US-0155139P.
PR	23-SEP-1999;	99US-0155486P.
PR	24-SEP-1999;	99US-0155653P.
PR	28-SEP-1999;	99US-0156458P.
PR	28-SEP-1999;	99US-0156596P.
PR	04-OCT-1999;	99US-0157117P.
PR	05-OCT-1999;	99US-015753P.
PR	06-OCT-1999;	99US-0157865P.
PR	07-OCT-1999;	99US-0158029P.
PR	08-OCT-1999;	99US-0158232P.
PR	12-OCT-1999;	99US-0158369P.
PR	13-OCT-1999;	99US-0158293P.
PR	13-OCT-1999;	99US-0158294P.
PR	13-OCT-1999;	99US-0158295P.
PR	14-OCT-1999;	99US-0159329P.
PR	14-OCT-1999;	99US-0159330P.
PR	14-OCT-1999;	99US-0159331P.
PR	14-OCT-1999;	99US-0159637P.
PR	14-OCT-1999;	99US-0159638P.
PR	18-OCT-1999;	99US-0159684P.
PR	21-OCT-1999;	99US-0160741D.
PR	21-OCT-1999;	99US-0160767P.
PR	21-OCT-1999;	99US-0160768P.
PR	21-OCT-1999;	99US-0160770P.
PR	21-OCT-1999;	99US-0160814P.
PR	21-OCT-1999;	99US-0160815P.
PR	22-OCT-1999;	99US-0160980P.
PR	22-OCT-1999;	99US-0160981P.
PR	22-OCT-1999;	99US-0160989P.
PR	23-OCT-1999;	99US-0161404P.
PR	23-OCT-1999;	99US-0161405P.
PR	23-OCT-1999;	99US-0161406P.
PR	26-OCT-1999;	99US-0161359P.
PR	26-OCT-1999;	99US-0161360P.
PR	28-OCT-1999;	99US-0161361P.
PR	28-OCT-1999;	99US-0161920P.
PR	28-OCT-1999;	99US-0161992P.
PR	28-OCT-1999;	99US-0161993P.
PR	29-OCT-1999;	99US-0161993P.
PR	29-OCT-1999;	99US-0162142P.
Alignment Scores:		
Pred. No.:	7.4e-44	Length: 1154
Score:	548.50	Matches: 125
Percent Similarity:	55.69%	Conservative: 61
Best Local Similarity:	37.43%	Mismatches: 125
Query Match:	26.64%	Indels: 23
DB:	3	Gaps: 10
US-09-854-122-16 (1-391) x AAC37672 (1-1154)		
Oy	22 LeuGLuLysCysPheGlySer	-----LySAsnGIuInGIuLYSGlu--- 35
Db	54 CTCTCTTTGTGTTTCAATCGGGACAAATTATCACATAGGAGGCTCTTAAGAAGCT 113	
Oy	36 -----GlusAspSerLysMetTyLysArgSTyrArgGluIleVal 48	
Db	114 CATCACTCCAAACTACATGAAGAACACACACGTAGTCAGAAACCAAGAACTTGATC 173	
Oy	49 SerSerLeuProSerAsn---AspTyrTrpGlyAspThrMetArgLeuTyLysGlyPhe 67	

```
Db 174 ACTTCTCTACCTTCACAAAGATTTCATGGGTTATGGTCTCTACAACTACAAAGGTTGT 233
Qy 68 TtpGlnMetGlyTyrLeuValProGlyIleMetAlaPheGluAspAsnPhelYsAlaArg 87
Db 234 TGGTACTATCCAAACACACTCCAAAGCGTCTTTCAGCTCCAAACACTCTTCAGGCCAGA 293
Qy 88 GluThrAspIleLeuThrThrLeuProLysAlaGlyThrThrThrThrThrThrThrThr 107
Db 294 GATCTGTATATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 353
Qy 108 ThrPheAlaIleLeuThrArgAspValAsnHisProSerSerProThrHisProLeuLeu 127
Db 354 ATTTCCTCTGTGTACATAGAAAGTACCGCGGACCCCTCAACACATCTCTTGTCTC 413
Qy 128 PhePheAsnProHisSerCysValGlnAsnLeuGluTyr---LeuTyrMetGlyArgGlu 146
Db 414 TTACAAACCCCTCATGACCTTGTCCCATTTCTTGGAGTTGAGTTATAGCTTAATAGCCAA 473
Qy 147 AsnThrMetProAspLeuAspMetLeuAsnGluSerProArgLeuPheAlaGlyHisIle 166
Db 474 -----ATTCCGATCTCCGAAGTATCT---TCTCTATGATCTTTTCTACACATG 524
Qy 167 ProTyrSerLeuProAlaSerValLeuLysSerGlyThrLysIleLeuAsnIleSer 186
Db 525 CACTTACAAGCATTTGGTGAAGCCACCAAAAGCT---TGCAAAACCCGTATATGTGT 581
Qy 187 ArgAsnArgLysSerThrPheValSerPheTyrPheGlyAsnLeuIleAsnProAsp 206
Db 582 AGAGGTATCAAGATACGTTTCTCCGCTGCATATAGAAACATGTTGATCGTCCACC 641
Qy 207 LysLeu-----LeuAspLeuGluLysSerValAspIlePheAlaSerGlyIleSerPhe 224
Db 642 AGATGATCAAGCCACTTTGAGCTCATGTTTGTATGCTTATTGTAGAGGAGTCTCTT 701
Qy 225 CysGlyProGluTyrAsnPheGlnAlaGluPhe---ThrAsnAlaIleSerThrAsnSer 243
Db 702 TATGGACCTTATTTGGGAACATGTAATGAGCTATTGGAAAGGGAGCTTGGAAAGGAG 761
Qy 244 AsnLeuLeuLeuLeuSerTyrGluGluMetLeuGluLysProValGluAsnValLysLys 263
Db 762 AATGTTCTTTTCATGAGTACGAGAGATTAATGAGGAGCTCGTTCCTCAAGTCAAGAGA 821
Qy 264 LeuAlaGluPheMetGlyCysGlyPheThrAspAspGluGluGlnGlyIleValAsp 283
Db 822 CTCGCGAGTCTTGGAAATGTCATTCACCAAGGAAGAAAGAAAGTGGATCGGTGGAG 881
Qy 284 GluIleValLysLeuCysSerPheAspAsnLeuLysAsnGlnGlnValAsnLysAsnGly 303
Db 882 GAGATCTTGAAGTTGTAGTTTACGAATTTAAGCAATTTGAGGATTTAAGATGGG 941
Qy 304 SerSerTyrAsnSerLysIleAspAsnLysHisPhePheArgLysGlyGluValArgAsp 323
Db 942 ACAAG---AGAATTGGTGTAGATTCTCAGGTCTCTTTAGGAAGAGTGAAGTTGTGAT 998
Qy 324 TtpAlaAsnTyrLeuThrSerGluMetIleLysLysLeuGlu 337
Db 999 TGAAGAATCATCTTACGCCAACAAATGGCGCAACCTTTGAT 1040

RESULT 14
ABZ13222
ID ABZ13222 standard; DNA; 996 BP.
XX AC ABZ13222;
XX DT
XX 21-JAN-2003 (first entry)
XX DE Arabidopsis thaliana stress regulated gene SEQ ID NO 1027.
XX KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX OS Arabidopsis thaliana.
XX
```

```
PN W0200216555-A2.
XX 28-FEB-2002.
XX 24-AUG-2001; 2001WO-US026685.
XX 24-AUG-2000; 2000US-0227866P.
XX 26-JAN-2001; 2001US-0264647P.
XX 22-JUN-2001; 2001US-0300111P.
XX (SCRI ) SCRIPPS RES INST.
XX PA (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX PI Harper JF, Krepes J, Wang X, Zhu T;
XX WPI; 2002-304127/34.
XX
XX Identifying a stress condition to which a plant cell has been exposed and
XX producing plants with increased tolerance to these abiotic stresses.
XX
XX Claim 144; SEQ ID NO 1027; 577bp + Sequence Listing; English.
XX
XX The invention relates to identifying a stress condition to which a plant
XX cell has been exposed, comprising: (a) contacting nucleic acid
XX representative of expressed polynucleotides in the plant cell with an
XX array or probes representative of the plant cell genome; and (b)
XX detecting a profile of expressed polynucleotides in the plant cell
XX characteristic of a stress response. The method is useful in the
XX production of transgenic plants, cells and seeds and in producing plants
XX with increased tolerance to abiotic stress. The present sequence is that
XX of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used
XX in methods of the invention. Note: The sequence data for this patent is
XX not represented in the printed specification but is based on sequence
XX information supplied to Derwent by the European Patent Office
XX
XX SQ Sequence 996 BP; 306 A; 199 C; 213 G; 278 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No.: 1,34e-43 Length: 996
XX Score: 545.00 Matches: 118
XX Percent Similarity: 58.31% Conservative: 61
XX Best Local Similarity: 38.44% Mismatches: 118
XX Query Match: 26.47% Indels: 10
XX DB: 6 Gaps: 8
XX
XX US-09-854-122-16 (1-391) x ABZ13222 (1-996)
XX
XX Qy 36 GluAspSerLysMetTyrLysArgTyrArgGluIleValSerSerLeuProSerAsn--- 54
XX Db 43 AAAGCAGCACACGTTAGTCAAGAAACCAAGAACTTGATCACTTCTCTACCTTCAGACAAA 102
XX Qy 55 AspTyrTrpGlyAspThrMetArgLeuTyrLysGlyPheTrpGlnMetGlyTyrLeuVal 74
XX Db 103 GATTTCATGGGTTATGGTCTCTCAACACTACAAAGGTTGGTGGTACTATCCAAACACACTC 162
XX Qy 75 ProGlyIleMetAlaPheGluAspAsnPhelYsAlaArgGluThrAspIleLeuThr 94
XX Db 153 CAAGCGGTCTTGAGTCCAAACACTTCAAGCCAGAGATCTGATATATATCTCGCT 222
XX Qy 95 ThrLeuProLysAlaGlyThrThrThrThrThrThrThrThrThrThrThrThrThrArg 114
XX Db 223 TCTTTGCCAAAGGTGGAACCACTTGGCTCAATCCCTTAATTTTCGTGTGTACATAGA 282
XX Qy 115 AspValAsnHisProSerSerProThrHisProLeuLeuPhePheAsnProHisSerCys 134
XX Db 283 GAAAGTACCCGGAAACCCCTCAACACATCTTTGCTCTTACAAAACCCCTCATGACCTT 342
XX Qy 135 ValGlnAsnLeuGluTyr---LeuTyrMetGlyArgGluAsnThrMetProAspLeuAsp 153
XX Db 343 GTCCCATTTCTTGAGTTGAGTTATAGCTTAATAGCCAA-----ATTCCGATCTCGCA 396
XX Qy 154 MetLeuAsnGluSerProArgLeuPheAlaGlyHisIleProTyrSerLeuLeuProAla 173
XX
```



D <sub>b</sub>	736	TTGGAAATGTGCTTTATTGAAGAAGAGAA-----CTGAGAGAGATTGTGAAG	783
Q <sub>y</sub>	288	LeuCysSerPheAspAsnLeuLysAsnGlnValAsnLysAsnGlySerTyrAsn	307
D <sub>b</sub>	784	TTGTGTAGCTTTGAGAGTTTAAAGTAAATTGCGAAGTTAACTAAGAGGGGAAATTCGCAAA	843
Q <sub>y</sub>	308	SerLysIleAspAsnLysHisPhePheArgLysGlyValArgAspTrpAlaIleTyr	327
D <sub>b</sub>	844	GGA---ATAGAGACTAAACACTTCTTTAGAAAGGAGAGATTGGAGAGATGAGAGATCT	900
Q <sub>y</sub>	328	LeuThrSerGluMetIleLysLysLeuGlu	337
D <sub>b</sub>	901	TTGAGTGAGTCATTGGCAGAGAAATTCAT	930

Search completed: March 20, 2004, 04:00:57  
Job time : 518 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 20, 2004, 03:46:04 ; Search time 97 Seconds  
(without alignments)  
2236.967 Million cell updates/sec

Title: US-09-854-122-16  
Perfect score: 2059  
Sequence: 1 TRGIGIAVACLELIMAGIL.....CLLSYFVNNFILLKKKK 391

Scoring table: BLOSUM62

	Xgapop 10.0	Xgapext 0.5
Xgapop 10.0	10.0	0.5
Fgapop 6.0	6.0	7.0
Delop 6.0	Delext 7.0	

Searched: 682709 segs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame\_p2n.model -DEV=xlh  
-Q/cgn2\_1/USPTO.spool/US09854122/runat\_18032004\_100001\_3276/app\_query.fasta\_1.583  
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NOEM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09854122.CCN\_1\_1\_56@runat\_18032004\_100001\_3276 -NCPU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA.\*

	1:	2:	3:	4:	5:	6:
/cgn2_6/prodata/2/ina/5A.COMB.seq.*						
/cgn2_6/prodata/2/ina/5B.COMB.seq.*						
/cgn2_6/prodata/2/ina/6A.COMB.seq.*						
/cgn2_6/prodata/2/ina/6B.COMB.seq.*						
/cgn2_6/prodata/2/ina/6C.COMB.seq.*						
/cgn2_6/prodata/2/ina/backfiles1.seq.*						

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	273	13.3	1063	1	US-08-325-562-1
2	273	13.3	1063	1	US-08-437-795-1
3	268	13.0	1004	4	US-09-609-816-3
4	264	12.8	942	4	US-09-609-816-1
5	260	12.6	922	4	US-09-609-816-2
6	239	11.6	798	4	US-09-795-926-5
7	239	11.6	912	4	US-09-795-926-1
8	228	11.1	1366	4	US-09-328-1748-2
9	209.5	10.2	2153	4	US-09-795-926-19
10	200	9.7	1564	2	US-08-852-481-1
11	187	9.1	1130	4	US-09-786-240-22
12	184	8.9	645	4	US-09-795-926-15

	13	155.5	7.6	447	4	US-09-795-926-7	Sequence 7, Appli
	14	155.5	7.6	561	4	US-09-795-926-11	Sequence 11, Appli
	15	124	6.0	3309	4	US-09-510-543-20	Sequence 20, Appli
	16	124	6.0	3614	4	US-09-023-655-968	Sequence 968, App
	17	124	6.0	5878	4	US-09-510-543-18	Sequence 18, Appli
	18	119	5.8	5181	1	US-08-357-073-10	Sequence 10, Appli
	19	117	5.7	3648	1	US-08-053-614-1	Sequence 1, Appli
	20	117	5.7	3648	2	US-08-316-397B-1	Sequence 1, Appli
	21	117	5.7	3648	2	US-09-034-306-1	Sequence 1, Appli
	22	117	5.7	3648	3	US-09-259-437-1	Sequence 1, Appli
	23	117	5.7	3648	5	PCT-US93-09782-1	Sequence 3, Appli
	24	117	5.7	4821	1	US-08-053-614-3	Sequence 3, Appli
	25	117	5.7	4821	1	US-08-316-397B-3	Sequence 3, Appli
	26	117	5.7	4821	2	US-09-034-306-3	Sequence 3, Appli
	27	117	5.7	4821	3	US-09-259-437-3	Sequence 3, Appli
	28	117	5.7	4821	5	PCT-US93-09782-3	Sequence 3, Appli
	29	111	5.4	333	4	US-09-795-926-3	Sequence 3, Appli
	30	111	5.4	447	4	US-09-795-926-9	Sequence 4, Appli
	31	111	5.4	5925	3	US-08-470-260-4	Sequence 4, Appli
	32	111	5.4	5925	3	US-08-471-491-4	Sequence 4, Appli
	33	111	5.4	5925	3	US-08-466-662-4	Sequence 4, Appli
	34	111	5.4	10299	2	US-08-477-451-1	Sequence 5, Appli
	35	111	5.4	10299	2	US-08-477-451-5	Sequence 5, Appli
	36	111	5.4	19932	2	US-08-477-451-25	Sequence 25, Appli
	37	111	5.4	640681	4	US-09-790-988-1	Sequence 1, Appli
	38	105	5.1	129908	4	US-09-585-858-1	Sequence 1, Appli
	39	104	5.1	41708	4	US-09-470-512A-3	Sequence 3, Appli
	40	103	5.0	4404	4	US-09-134-000C-3335	Sequence 3335, Ap
	41	103	5.0	11015	4	US-10-204-708-55	Sequence 55, Appli
	42	102	5.0	580073	4	US-08-545-528B-1	Sequence 1, Appli
	43	100.5	4.9	294	4	US-09-795-926-17	Sequence 17, Appli
	44	99.5	4.8	3385	4	US-09-509-800-1	Sequence 1, Appli
	45	99.5	4.8	9636	1	US-08-323-170B-1	Sequence 1, Appli

#### ALIGNMENTS

RESULT 1  
US-08-325-562-1  
; Sequence 1, Application US/08325562  
; Patent No. 5714594  
; GENERAL INFORMATION:  
; APPLICANT: Weinshilboum, Richard M.  
; APPLICANT: Aksoy, Ibrahim A.  
; APPLICANT: Wood, Thomas C.  
; TITLE OF INVENTION: CDNA CLONING AND EXPRESSION OF HUMAN  
; TITLE OF INVENTION: LIVER ESTROGEN SULFOTRANSFERASE  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Schwegman, Lundberg & Woessner P.A.  
; STREET: 3500 IDS Center  
; CITY: Minneapolis  
; STATE: MN  
; COUNTRY: USA  
; ZIP: 55402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/325,562  
; FILING DATE: 18-OCT-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mueeting, Ann M.  
; REGISTRATION NUMBER: 33,977  
; REFERENCE/DOCKET NUMBER: 150.131US1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 612-339-0331  
; TELEFAX: 612-339-3061  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:

```

; LENGTH: 1063 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 107..989
US-08-325-562-1

Alignment Scores:
Pred. No.: 7,06e-23 Length: 1063
Score: 273.00 Matches: 99
Percent Similarity: 42.62% Mismatches: 140
Best local Similarity: 27.05% Gaps: 16
Query Match: 13.26%
DB: 1

US-09-854-122-16 (1-391) x US-08-325-562-1 (1-1063)

QY 41 TyrLysArgTyrArgGluLeuValSerSerLeuProSerAsnAspTyrTyrGlyAspThr 60
DB 128 TATGAAAGATTGAGAGATGCTCATGGGATCTA-----160
QY 61 MetArgLeuTyrLysGlyPheTyrPgluMetGlyTyrLeuValProGlyLeuMetAlaPhe 80
DB 161 -----ATGTATAAGATTT-----GTCAATAT 184
QY 81 GluAspAsn-----PheValaArgGluThrAspLeuLeuLeuThrLeuPro 97
DB 185 TGGGATATGGAAGCGTTCCAGCAAGACCAATGATCTGTGTCATGTCACCTACCT 244
QY 98 LysAlaGlyThrThrTyrThrLysAlaLeuThrPheAlaLeuThrArg-----AspVal 116
DB 245 AAATCTGTACAACTGCTGTTAGTCAATCTGTATATGATCTATAAAGAGGGTGTGTG 304
QY 117 AsnHisProSerSerProThrHisProLeuLeuPhePheAsnProHisSerCysValGln 136
DB 305 GAAAGTGCAGAAAGAT-----GTAATTTTAAAT-----334
QY 137 AsnLeuGluTyrLeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsn 156
DB 335 CGAATACCTTCTCGAATGCGAGAAAGAAACCTCATGATGGAGTAAACAAATAGAT 394
QY 157 Glu-----SerProArgLeuPheAlaGlyHisLeuProTyrSerLeuLeuProAlaSer 174
DB 395 GAGATCAATCTCTAGAACTGAGACTCATTTGCCACCTGAACTCTTCTCCCTCCCA 454
QY 175 ValLeuLysSerGlyThrLysLeuLeuLeuSerArgAsnArgLysSerThrPheVal 194
DB 455 TTTTGGGAAAGGATGTAGATATCTATCTTCCCGGAATGCAAGGATGTGCTGT 514
QY 195 SerPheTyrLysPheGlyAsnLeuLeu-----AsnProAspLysLeuLeuAsp 210
DB 515 TCCTTTTATTATCTCTTCTAATGTTGGTGTCTATCTTCCGGAATCTCGGATCTTCCAGAG 574
QY 211 LeuGluLysSerValAspLeuPheAlaSerGlyLeuSerPheCysGlyProGluTyrAsn 230
DB 575 TTT-----GTGAGAAATTCATGCAAGACAGGTTCTTATGTT-----TCCTGGAT 622
QY 231 PheGlnAlaGluPheThrAsnAlaLaSerThrAsnSerAsnLeuLeuLeuSerTyr 250
DB 623 AAACATGTAATAATCTTGTGGGAAAGGAAAGATCCACGTGACTATCTTTTCTTAC 682
QY 251 GluGluMetLeuGluLysProValGluAsnValLysLeuAlaGluPheMetGlyCys 270
DB 683 GAAGACCTGAAGAGATATCAGAAAGAGGATGATTAATGATACATTTCTCTG-----736
QY 271 GlyPheThrAspAspGluLysGlnGlyLeuValAspGluLeuValLysLeuCysSer 290
DB 737 -----GAAAGGAAGCCATCAGAGGAGCTGTGACAGGATATATACATCATCTCTCG 787
QY 291 PheAspAsnLeuLysAsnGlnGlnValAsnLysAsnGlySerSerTyrAsnSerLysile 310

```

```

DB 788 TTCCAAGAGATGAAGAACAATCCA---TCCAAATATTACACACACTGCCAGCAAT 844
QY 311 AsphenLysHis-----PhePheArgLysGlyGluValArgAspTyrAlaAsnTyr 327
DB 845 ATGAACCAAGAAATGTGCCCTTTCATGAGAAGGGAATTCAGGAGACTGGAAAAATCAC 904
QY 328 LeuThrSerGluMetIleLysLysLeuGluThrAlaGlyLysIleAsnGluSerGluLys 347
DB 905 TTTACAGTAGCCCTGAATGAAAAATTT-----GATAAA 937
QY 348 HisLeuLeuSerAsnLysAsnLeuThrAsnPheAsnLeuAsnAsnTyrCysGluLysSer 367
DB 938 CATTTAGCAGCAAAATGAAGAAATCTACACTGAAGTTT-----CGAAT 982
QY 368 AsnIle-----SerLeuCysLeuLeuSerTyrSerPheValIleAsnAsnPheIleLeu 385
DB 983 GAGATCTAAGAAGTCTTTCTTAAACATATCTGATATTAAGATTTCTTTCTTCA 1042
QY 386 LeuLysLysLysLysLysLys 391
DB 1043 TTCAAAAAAATAAAAAA 1060

RESULT 2
US-08-437-795-1
; Sequence 1, Application US/08437795
; Patent No. 5744355
; GENERAL INFORMATION:
; APPLICANT: Weinshilbom, Richard M.
; APPLICANT: Kinsley, Ibrahim A.
; APPLICANT: Wood, Thomas C.
; TITLE OF INVENTION: CDNA CLONING AND EXPRESSION OF HUMAN
; TITLE OF INVENTION: LIVER ESTROGEN SULFOTRANSFERASE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg & Woessner P.A.
; STREET: 3500 IDS Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/437,795
; FILING DATE: 09-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mueeting, Ann M.
; REGISTRATION NUMBER: 33,977
; REFERENCE/DOCKET NUMBER: 150.137US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-3061
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1063 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 107..989
US-08-437-795-1

Alignment Scores:
Pred. No.: 7,06e-23 Length: 1063
Score: 273.00 Matches: 99
Percent Similarity: 42.62% Mismatches: 140
Query Match: 13.26%
DB: 1

```

Best Local Similarity: 27.05% Mismatches: 140  
Query Match: 13.26% Indels: 70  
DB: 1 Gaps: 16

US-09-854-122-16 (1-391) x US-08-437-795-1 (1-1063)

QY 41 TyrLysArgTyrArgGluIleValSerSerLeuProSerAsnAspTyrTrpGlyAspThr 60  
DB 128 TATGAAGATTGGAAGAGTCCATGGATTCTA----- 160  
QY 61 MetArgLeuTyrLysGlyPheTrpGlnMetGlyTyrLeuValProGlyIleMetAlaPhe 80  
DB 161 -----ATGTATAAAGATTT-----GTCAATAT 184  
QY 81 GluAspAsn-----PheLysAlaArgGluThrAspIleLeuThrLeuPro 97  
DB 185 TGGGATATGTGAAGCGTTCAGGACAGCAGATGATGTGTCATTCGCCACCTACCT 244  
QY 98 LysAlaGlyThrTrpThrLysAlaLeuThrPheAlaLeuThrArg---AspVal 116  
DB 245 AAATCTGGTACAACTGGGTAGTGAATTTGTATATGATCTATAAAGAGGGTATGTG 304  
QY 117 AsnHisProSerSerProThrHisProLeuLeuPheAsnProHisSerCysValGln 136  
DB 305 GAAAGTGCAGAGAGAT-----GTAATTTTAAT----- 334  
QY 137 AsnLeuGluTyrLeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsn 156  
DB 335 CGAATACCTTTCTCGAATGCAGAAAGAAACCTCATGATGGAGTAAACAAATAGAT 394  
QY 157 Glu-----SerProArgLeuPheAlaGlyHisIleProTyrSerLeuProAlaSer 174  
DB 395 GAGATGAATTCCTAGAAITGTGAAGACTAATGGCCACCTGAACCTTCCTCCCTCA 454  
QY 175 ValLeuLysSerGlyThrLysIleLeuAsnIleSerArgAsnArgLysSerThrPheVal 194  
DB 455 TTTTGGGAAAAGGATTGTAAGATAATCTATCTTTGCCGGAATGCAAGGATGGCTGT 514  
QY 195 SerPheTrpLysPheGlyAsnLeuLe-----AsnProAspLysLeuLeuAsp 210  
DB 515 TCTTTTATATTCTTCTATATGGTGGTGTGTCATCCAAATCCTGGATCCTTTCCAGAG 574  
QY 211 LeuGluLysSerValAspIlePheAlaSerGlyIleSerPheCysGlyProGluTrpAsn 230  
DB 575 TTT-----GTGGAGAAATTCATGCAAGACAGAGTTCTTATGTT--TCCTGGTAT 622  
QY 231 PheGluAlaGluPheThrAsnAlaAlaSerThrAsnSerAsnLeuLeuLeuSerTyr 250  
DB 623 AAACATGTATAAATCTTGGTGGGAAAAGGAGAGTCCACGTGACTATTTCTTTCTAC 682  
QY 251 GluGluMetLeuGluLysProValGluAsnValLysLeuAlaGluPheMetGlyCys 270  
DB 683 GAAGACTGAAGAGGATATCAGAAAGAGGTGATAAATGTATACATTTCTCTG----- 736  
QY 271 GlyPheThrAspAspGluGluLysGlnGlyIleValAspGluIleValLysLeuCysSer 290  
DB 737 -----GAAAGGAAGCCATCAGAGAGCTTGTGGACAGGATTATACATCTCTCG 787  
QY 291 PheAspAsnLeuLysAsnGlnValAsnLysAsnGlySerSerTyrAsnSerLysIle 310  
DB 788 TTCAGAGATGAAGAAATCA-----TCCAAATATACACACTGCCAGCAAT 844  
QY 311 AspAsnLysHis-----PhePheArgLysGlyGluValArgAspTrpAlaSer 327  
DB 845 ATGACCAGAAATTTGCGCCCTTCATGAGAAGGAATTACAGGACTGGAATAATCAC 904  
QY 328 LeuThrSerGluMetIleLysLeuGluThrAlaGlyLysIleAsnGluSerGluLys 347  
DB 905 TTTACAGTACCTCGAATGAAATTT-----GATAAA 937  
QY 348 HisLeuLeuSerAsnLysAsnLeuThrAsnPheAsnLeuAsnTyrCysGluLysSer 367  
DB 938 CATTTAGCAGCAATGAAGGAATCTACACTGAAGTTT-----CGAATC 982

368 AsnIle-----SerLeuCysLeuLeuSerTyrSerPheValIleAsnPheIleLeu 385  
DB 983 GAGATCTAAGAGCTCTTTCTTTACTTTACATATCTGATATTAAAGATTCTTTTCATTA 1042  
QY 386 LeuLysLysLysLysLys 391  
DB 1043 TTTCAAAAAAATAAAAAA 1060

RESULT 3  
US-09-609-816-3  
; Sequence 3, Application US/09609816  
; Patent No. 643684  
; GENERAL INFORMATION:  
; APPLICANT: Woodage, Trevor  
; APPLICANT: Wei, Minh Hui  
; APPLICANT: Kodira, Chinappa  
; APPLICANT: Beasley, Ellen  
; APPLICANT: DiFrancesco, Valentina  
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING  
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN  
; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF  
; FILE REFERENCE: CL000669PCT  
; CURRENT APPLICATION NUMBER: US/09/609,816  
; CURRENT FILING DATE: 2000-07-03  
; PRIOR APPLICATION NUMBER: 60/192,408  
; PRIOR FILING DATE: 2000-03-27  
; PRIOR APPLICATION NUMBER: 60/212,725  
; PRIOR FILING DATE: 2000-06-20  
; PRIOR APPLICATION NUMBER: 09/609,816  
; PRIOR FILING DATE: 2000-07-03  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 1004  
; TYPE: DNA  
; ORGANISM: HUMAN  
US-09-609-816-3

Alignment Scores:  
Pred. No.: 2,568-22 Length: 1004  
Score: 268.00 Matches: 97  
Percent Similarity: 43.97% Conservative: 56  
Best Local Similarity: 27.87% Mismatches: 131  
Query Match: 13.02% Indels: 64  
DB: 4 Gaps: 18

US-09-854-122-16 (1-391) x US-09-609-816-3 (1-1004)

QY 29 LysAsnGluGlnGluLysGluAspSerLysMetTyrLysArgTyrArgGluIleVal 48  
DB 62 AAAAAACGCTCCACGATCGAAGAAAAAGCCAGAACTGTTTAAACATCATGGA--GTA 115  
QY 49 SerSerLeuProSerAsnAspTyrTrpGlyAspThrMetArgLeuTyrLysGlyPheTrp 68  
DB 116 GATGGAGTCCCT-----ACGTGATATTATCAAAAGATGTGG 154  
QY 69 GlnMetGlyTyrLeuValProGlyIleMetAlaPheGluAspAsnPheLysAlaArgGlu 88  
DB 155 GAAAAA-----GTCTGTAATTTCCAAAGCCAGCCT 184  
QY 89 ThrAspIleLeuThrThrLeuProLysAlaGlyThrTrpThrLysAlaLeuThr 108  
DB 185 GATGATCTTATCTGGCACTTACCAAGAGTCAGGTACAAACATGGATGCATGAATTTA 244  
QY 109 PheAlaIleLeuThr---ArgAspValAsnHis-----ProSerSerProThr 123  
DB 245 GACATGATCTTAATGATGGTGAATGTGAGAAATGCAAAAGAGCCAGACTCTAGATAGA 304  
QY 124 HisProLeuLeuPhePheAsn---ProHisSerCysValGlnAsnLeuGluTyrLeuTyr 142  
DB 305 CACGCTTCTTGAACCTGAAATTTCCCATTAAGAAAAACAGATTTGGAGTTTCGTT--- 361



```
Db 746 GCCAACCATCTGCGGTACTGCTGCATATTCATCACTCCATCTCATAAATTTATGAGG 805
Qy 318 LysGluValArgAspTrpAlaAsnTyrLeuThrSerGluMetIleLysLysLeuGlu 337
Db 806 AAAGGGATGCTGAGACTGGAAGAACCACTTTACTGTG-----844
Qy 338 ThrAlaGlyLysIleAsnGluSer---GluLysHisLeuLeuSerAsnLysAsnLeuThr 356
Db 845 -----GCTTGTGATGGAACCTTTGATAGCACTAT-----GAAAGAAGATGGCA 889
Qy 357 AsnPheAsnLeuAsnAsnTyrCys 364
Db 890 GGGTCCACACTG---AACTTCTGC 910

RESULT 5
US-09-609-816-2
; Sequence 2, Application US/09609816
; Patent No. 643684
; GENERAL INFORMATION:
; APPLICANT: Woodage, Trevor
; APPLICANT: Wei, Minh Hui
; APPLICANT: Kodira, Chinappa
; APPLICANT: Beasley, Ellen
; APPLICANT: Difrancesco, Valentina
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL000669PCT
; CURRENT APPLICATION NUMBER: US/09/609,816
; CURRENT FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/192,408
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: 60/212,725
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/609,816
; PRIOR FILING DATE: 2000-07-03
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 922
; TYPE: DNA
; ORGANISM: HUMAN
US-09-609-816-2

Alignment Scores:
Pred. No.: 2,03e-21 Length: 922
Score: 260.00 Matches: 86
Percent Similarity: 45.02% Conservative: 54
Best Local Similarity: 27.65% Mismatches: 123
Query Match: 12.63% Indels: 48
DB: 4 Gaps: 13

US-09-854-122-16 (1-391) x US-09-609-816-2 (1-922)
Qy 29 LysAsnGluGlnGluLysGluGluAspSerLysMetTyrLysArgTyrArgGluIleVal 48
Db 23 AAAAAAGCTCCACGATGGAAGAAAAAGCCAGAACTGTTAAATCATCATGGAA-----GTA 76
Qy 49 SerSerLeuProSerAsnAspTyrTyrGlyAspThrMetArgLeuTyrLysGlyPheTyr 68
Db 77 GATGAGTGCCT-----ACGTTGATATATCAAAAGAAATGGTGG 115
Qy 69 GlnMetGlyTyrLeuValProGlyIleMetAlaPheGluAspAsnPhelysAlaArgGlu 88
Db 116 GAAAA-----GTATGTAATTTCCAAGCAAGCCT 145
Qy 89 ThrAspIleLeuThrThrLeuProLysAlaGlyThrThrThrThrLysAlaLeuThr 108
Db 146 GATGATCTTATCTGGCACTTACCCAAAGTCAGGTACACATGGATGCATGAATTTTA 205
Qy 109 PheAlaIleLeuThr---ArgAspValAsnHis-----ProSerProThr 123
Db 206 GACATGATTTCTAATGATGGTGTGATGTGAGAAATGTCGAAAGAGCCAGCACTCTAGATGA 265
```

```
Qy 124 HisProLeuLeuPhePheAsn---ProHisSerCysValGlnAsnLeuGluTyrLeuTyr 142
Db 266 CACCTTTCCTTGGACTGAAATTTCCCATTAAGAAAAACACAGATTGGAGTTGGT---322
Qy 143 MetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsnGluSerProArgLeuPhe 162
Db 323 -----CTTCAATG-----TCCTCACCACCACTGATA 349
Qy 163 AlaGlyHisIleProTyrSerLeuLeuProAlaSerValLeuLysSerGlyThrLysIle 182
Db 350 AAAACACATCTCCCTTCATCTCATCTGATTCACCATCTATCTGGAAGAAAACTGCAAGATT 409
Qy 183 IleAsnIleSerArgAsnArgLysSerThrPheValSerPheThrLysPheGlyAsnLeu 202
Db 410 GTCTATGTGGCCAGAAATCCAGGATTGCTGTGTCTACTACTACCACTTTACAGGATG 469
Qy 203 IleAsn-----ProAspLysLeuLeuAspLeuLysSerValaAspIlePheAlaSer 220
Db 470 GCTTCCTTTATGCTGCTGATCCT---CAGAACCTTAGAGGAATTTATGAGAAATTCATGTCC 526
Qy 221 GlyIleSerPheCysGlyProGluTrpAsnPheGlnAlaGluPheThrAsnAlaAlaSer 240
Db 527 GGA---AAAGTTGTTGGGGCTCCTGTTGACCATGTGAAAGGATGCTGGCTGCAAAA 583
Qy 241 ThrAsnSerAsnLeuLeuLeuSerTyrGluGluMetLeuGluLysProValGluAsn 260
Db 584 GACATGCACCGGATCCTCTACCTCTCTACGAGGATATTAAGAAAGACCCAAAGCGGAA 643
Qy 261 ValLysLysLeuAlaGluPheMetGlyCysGlyPheThrAspAspGluGluLysGlnGly 280
Db 644 ATTGAGAAGATCTAGAACTTCTG-----GAAAAAGACATATCAGAGGAA 688
Qy 281 IleValAspGluIleValLysLeuCysSerPheAspAsnLeuLysAsnGlnValAsn 300
Db 689 ATTCTGATATAAATCATCTATCACACCTCCTTGTATGATGAAGCAAAACCCCAATGACC 748
Qy 301 LysAsnGlySerSerTyrAsnSerLysIleAspAsn-----LysHisPhePheArgLys 318
Db 749 AACTATACCACTTTGCCCAACGAGATTTATGACCACTCCCTCCCTTTTATGAGGAAA 808
Qy 319 GlyGluValArgAspTrpAlaAsnTyrLeuThr 329
Db 809 GGGATGCTGGAGACTGGAAGAACTATTTTACT 841

RESULT 6
US-09-795-926-5
; Sequence 5, Application US/09795926
; Patent No. 6555669
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kieke, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; PRIOR FILING DATE: 2000-03-24
```

; NUMBER OF SEQ ID NOS: 47  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 5  
 ; LENGTH: 798  
 ; TYPE: DNA  
 ; ORGANISM: homo sapiens  
 US-09-795-926-5

Alignment Scores:  
 Pred. No.: 5,29e-19 Length: 798  
 Score: 239.00 Matches: 79  
 Percent Similarity: 42.81% Conservativeness: 46  
 Best Local Similarity: 27.05% Mismatches: 121  
 Query Match: 11.61% Indels: 46  
 DB: 12 Gaps: 12

US-09-854-122-16 (1-391) x US-09-795-926-5 (1-798)

QY 82 AspAsnPhelYsAlaArgGluThrAspIleLeuThrLeuProLysAlaGlyThr 101  
 Db 31 GACACCTTGAAGCCAGACATGATGACATCGCTAGCATCTATCCAAAGTGGGTCA 90  
 QY 102 ThrTrp-----ThrLysAlaLeuThrPheAlaLeuThrArgAspValAsn 117  
 Db 91 AACTGGATTCTCCACATTCGATGCAATTAATATATGCTGTTCTTAAAAAAAGTATAAA 150  
 QY 118 HisProSerSerProThrHisProLeuLeuPheAsnProHisSerCysValGlnAsn 137  
 Db 151 TATCCAGATTCCACGTT-----CITGATGTTGGGGATTCA 186  
 QY 138 LeuGluThrLeuThrMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsnGlu 157  
 Db 187 GAAAAATAT-----CAGAGAATGAAAGCTTTTCCA----- 216  
 QY 158 SerProArgLeuPheAlaGlyHisIleProThrLeuLeuProAlaSerValLeuLys 177  
 Db 217 TCACCAAGGATTTCGCAACTCACCTCCATGACAAATACCTGGTCTATCTTCGAG 276  
 QY 178 SerGlyThrLysIleLeuAsnIleSerArgAsnArgLysSerThrPheValSerPheTrp 197  
 Db 277 AATAAAGCCCAAGATATGTTGATATTTCGAAACCCCTTAAAGATACAGCAGTATCTTTT 336  
 QY 198 LysPheGlyAsnLeuIleAsnProAspLysLeuLeuAspLeuGluLysSerValAspIle 217  
 Db 337 CATTTCCACACAGATGTC-----CCCGAT-----ATTCCAAGCTATGGCTCTGGATGAA 387  
 QY 218 PheAlaSerGlyIleSerPheCysGlyProGluThrAsnPheGlnAlaGluPheThr--- 236  
 Db 388 TTCTTCAGACAGTTCATGAAAGCAAGATTCTTGGGAAGGTATTTGATTTTGCATC 447  
 QY 237 -----AsnAlaAlaSerThrAsnSerAsnLeuLeuLeuLeuSerTyrGluGluMetLeu 254  
 Db 448 AATTGGAACAAACATCTTCATGCGCAGCAATGTTAAGTTCAATATATATGAAGACCTGAAA 507  
 QY 255 GluLysProValGluAsnValLysLeuAlaGluPheMetGlyCysGlyPheThrAsp 274  
 Db 508 GAGAATCTGCTCTCGAATAAAACAGATTGCTGAGTCTTGGGATTTCTTAACCTGGG 567  
 QY 275 AspGluGluLysGlnGlyIleValAspGluIleValLysLeuLysSerPheAspAsnLeu 294  
 Db 568 GAGCAA-----ATTCAAACATACTCAGTCCAGCAGCAGCCTTCCAGGCCATG 612  
 QY 295 LysAsnGlnGlnValAsnLysAsnGlySerSerTyrAsnSerLysIleAspAsnLysHis 314  
 Db 613 CGTGGAGAGTCTCAGGACACACACGCTGCTGTC-----GGCCCATTC 654  
 QY 315 PhePheArgLysGluValArgAspTrpAlaAsnTyrLeuThrSerGluMetIleLys 334  
 Db 655 CTTTTCGCAAGGTCAGTTCGTGATTCGAAAAAT---TTGTTTCAGTGAAATTCAGAAC 711  
 QY 335 LysLeuGluThrAlaGlyLysIleAsnGluSerGluLysHisLeuLeuSerAsnLysAsn 354  
 Db 712 CAG-----GAAATGGATGAAAAAATTCAAAGAGTGCTTAGCAGGCACCTCC 756

QY 355 Leu---ThrAsnPhelAsnLeuAsnAsnTyrCysGlu 365  
 Db 757 CTCGGAGCAAAAGTTGAGTAGTATGATATGTCGAG 792

# RESULT 7

US-09-795-926-1  
 ; Sequence 1, Application US/09795926  
 ; Patent No. 6555669  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Donoho, Gregory  
 ; APPLICANT: Hilbun, Erin  
 ; APPLICANT: Turner, C. Alexander Jr.  
 ; APPLICANT: Friedrich, Glenn  
 ; APPLICANT: Abuin, Alejandro  
 ; APPLICANT: Zambrowicz, Brian  
 ; APPLICANT: Sands, Arthur T.  
 ; APPLICANT: Walke, D. Wade  
 ; APPLICANT: Wilganowski, Nathaniel L.  
 ; APPLICANT: Hu, Yi  
 ; APPLICANT: Kieke, James Alvin  
 ; APPLICANT: Potter, David George  
 ; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
 ; FILE REFERENCE: LEX-0144-USA  
 ; CURRENT APPLICATION NUMBER: US/09/795,926  
 ; CURRENT FILING DATE: 2001-02-28  
 ; PRIOR APPLICATION NUMBER: US 60/185,920  
 ; PRIOR FILING DATE: 2000-02-29  
 ; PRIOR APPLICATION NUMBER: US 60/186,558  
 ; PRIOR FILING DATE: 2000-03-02  
 ; PRIOR APPLICATION NUMBER: US 60/191,849  
 ; PRIOR FILING DATE: 2000-03-24  
 ; NUMBER OF SEQ ID NOS: 47  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 1  
 ; LENGTH: 912  
 ; TYPE: DNA  
 ; ORGANISM: homo sapiens  
 US-09-795-926-1

## Alignment Scores:

Pred. No.: 6,55e-19 Length: 912  
 Score: 239.00 Matches: 79  
 Percent Similarity: 42.81% Conservativeness: 46  
 Best Local Similarity: 27.05% Mismatches: 121  
 Query Match: 11.61% Indels: 46  
 DB: 12 Gaps: 12

US-09-854-122-16 (1-391) x US-09-795-926-1 (1-912)

QY 82 AspAsnPhelYsAlaArgGluThrAspIleLeuThrLeuProLysAlaGlyThr 101  
 Db 145 GACACCTTGAAGCCAGACATGATGACATCGCTAGCATCTATCCAAAGTGGGTCA 204  
 QY 102 ThrTrp-----ThrLysAlaLeuThrPheAlaLeuThrArgAspValAsn 117  
 Db 205 AACTGGATTCTCCACATTCGATGCAATTAATATATGCTGTTCTTAAAAAAAGTATAAA 264  
 QY 118 HisProSerSerProThrHisProLeuLeuPheAsnProHisSerCysValGlnAsn 137  
 Db 265 TATCCAGATTCCACGTT-----CITGATGTTGGGGATTCA 300  
 QY 138 LeuGluThrLeuThrMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsnGlu 157  
 Db 301 GAAAAATAT-----CAGAGAATGAAAGCTTTTCCA----- 330  
 QY 158 SerProArgLeuPheAlaGlyHisIleProThrLeuLeuProAlaSerValLeuLys 177  
 Db 331 TCACCAAGGATTTCGCAACTCACCTCCATGACAAATACCTGGTCTATCTTCGAG 390  
 QY 178 SerGlyThrLysIleLeuAsnIleSerArgAsnArgLysSerThrPheValSerPheTrp 197



US-09-795-926-19  
; Sequence 19, Application US/09795926  
; Patent No. 6555669  
; GENERAL INFORMATION:  
; APPLICANT: Doncho, Gregory  
; APPLICANT: Hilbun, Erin  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Friedrich, Glenn  
; APPLICANT: Abuin, Alejandro  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sands, Arthur T.  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Wilganowski, Nathaniel L.  
; APPLICANT: Hu, Yi  
; APPLICANT: Kieke, James Alvin  
; APPLICANT: Potter, David George  
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
; FILE REFERENCE: LEX-0144-USA  
; CURRENT APPLICATION NUMBER: US/09/795,926  
; PRIOR FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: US 60/185,920  
; PRIOR FILING DATE: 2000-02-29  
; PRIOR APPLICATION NUMBER: US 60/186,558  
; PRIOR FILING DATE: 2000-03-02  
; PRIOR APPLICATION NUMBER: US 60/191,849  
; PRIOR FILING DATE: 2000-03-24  
; NUMBER OF SEQ ID NOS: 47  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 19  
; LENGTH: 2153  
; TYPE: DNA  
; ORGANISM: homo sapiens  
US-09-795-926-19

Alignment Scores:  
Pred. No.: 8,88e-15 Length: 2153  
Score: 209.50 Matches: 82  
Percent Similarity: 37.04% Conservative: 48  
Best Local Similarity: 23.36% Mismatches: 123  
Query Match: 10.17% Indels: 98  
DB: 4 Gaps: 13

US-09-854-122-16 (1-391) x US-09-795-926-19 (1-2153)

QY 82 AspAsnPheLysAlaArgGluThrAspIleLeuThrLeuProLysAlaGlyThr 101  
DB 678 GACACCTTCGAAGCCAGACATGATGACATCGCTGCTAGCATCTTATCCAAAGTGGGTCA 737  
QY 102 ThrTrp-----ThrLysAlaLeuThrPheAlaIleLeuThrArgAspValAsn 117  
DB 738 AACTGGATTCTCCACATTGTCAGTGAATTAATATATGCTGTTCTTAAAMAAAGTATAAA 797  
QY 118 HisProSerPro----- 122  
DB 798 TATCCAGAAATCCAGTCTTGAATGTGGGATTCCAGAAATATCTAGTGACAGTAGTG 857  
QY 123 -----ThrHisProLeuLeu 127  
DB 858 CTAGAAACTATCATCTAGATACCAAACTAGGAGTGATTCAACACACACACACA----- 911  
QY 128 PhePheAsnProHisSerCysValGlnAsnLeuGluTyrLeuTyr-MetGlyArgGluAs 147  
DB 912 -----CATGACACACAGTGGACATTTCACCTTTTGTGTATATTTTAAGAGAAATGAA 965  
QY 147 nThrMetProAspLeuAspMetLeuAsnGluSerProArgLeuPheAlaGlyHisIlePr 167  
DB 966 AGSCTTTCCA-----TCACCAAGGATTTGGCAACTCACCTCCA 1004  
QY 167 oTyrSerLeuLeuProAlaSerValIleuLysSerGlyThrLys----- 181  
DB 1005 CTATGACAAATTAACCTGGGTCTATCTTGGAGATTAAGCCAGACAGCATCTCATAT 1064

QY 181 ----- 181  
DB 1065 GTTGCACAGGCTGCTCGAACTCCTCAAGAGATCCTTCTGCCACCAAGGCTCC 1124  
QY 182 -----IleIleAsnIleSerArgAsnArgLysSerThrPheValSerPheTrpLysPh 199  
DB 1125 CAAAGTGATATTGGTGATATTTCGAAACCTCAAGATACAGACAGTATCTTTTTCGATT 1184  
QY 199 eGlyAsnLeuIleAsnProAspLysLeuLeuAspLeuGluLysSerValAspIlePheAl 219  
DB 1185 CCACACAGATGTC---CCCGAT-----ATTCCAGCTATGGCTCTTGGGATGAATCTT 1235  
QY 219 aSerGlyLysSerPheCysGlyProGluTrpAsnPheGlnAlaGluPheThr----- 236  
DB 1236 CAGACAGTTTCATGAAGACCAAGTTTCTTGGGAGAGTATTTGTTGCAATCAATTG 1295  
QY 237 -AsnAlaAlaSerThrAsnSerAsnLeuLeuLeuSerTyrGluGluMetLeuGluLys 256  
DB 1296 GAACAAACATCTTGCATGCGACATGTTAAAGTTTCATATATGAAGACCTGAAGAGAA 1355  
QY 256 sProValGluAsnValLysLysLeuAlaGluPheMetGlyCysGlyPheThrAspAspG 276  
DB 1356 TCTGCTGCTGGAATAAACAAGATGCTGAGTTCTTGGGATTTCTTCTACTCTGGGAGCA 1415  
QY 276 uGluLysGlnGlyIleValAspGluIleValLysLeuCysSerPheAspAsnLeuLysAs 296  
DB 1416 A-----ATTCAAACTATCTCAGTCCAGACACCTTCCAGCCATGGGTGC 1460  
QY 296 nGlnGluValAsnLysAsnGlySerSerTyrAsnSerLysIleAspAsnLysHisPhePh 316  
DB 1461 GAAGTCTCAGGACACACACGGTGTCTGC-----GGCCATCTCTTTT 1502  
QY 316 eArgLysGlyGluValArgAspTrpAlaAsnTyrLeuThrSerGluMetIleLysLysLe 336  
DB 1503 CCGCAAGGTGAGTGTGCTGATTGAAAAAT---TTGTTCAAGTAAATTCAGAACCCAG-- 1557  
QY 336 uGluThrAlaGlyLysIleAsnGluSerGluLysHisLeuLeuSerAsnLysAsnLeu-- 355  
DB 1558 -----GAAATGGATGAAATAATTCAAAGAGTGCTTAGCAGGCACCTCTCTCGG 1504  
QY 356 -ThrAsnPheAsnLeuAsnAsnTyrCysGlu 365  
DB 1605 AGCAAGTTGAAGTATGAATCATATTCATATTCGCAG 1635

RESULT 10  
US-08-852-481-1  
; Sequence 1, Application US/08852481  
; Patent No. 5928931  
; GENERAL INFORMATION:  
; APPLICANT: Grun Ph.D., Felix  
; APPLICANT: Buck Ph.D., Jochen  
; APPLICANT: Hammerling Ph.D., Ulrich  
; TITLE OF INVENTION: ISOLATION, PURIFICATION AND CLONING OF  
; TITLE OF INVENTION: RETINOL DEHYDRATASE  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Nixon, Hargrave, Devans & Doyle LLP  
; STREET: Clinton Square, P.O. Box 1051  
; CITY: Rochester  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 14603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/852,481  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/017,178







```

Db      604 CTCGGAGCAAGTGAAGTATGATCATATTCGCCAG 639

RESULT 13
US-09-795-926-7
; Sequence 7, Application US/09795926
; Patent No. 6555669
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kieke, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 447
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-795-926-7

Alignment Scores:
Pred. No.:      2,12e-09      Length:      447
Score:          155.50      Matches:      42
Percent Similarity: 45.74%      Conservative: 17
Best Local Similarity: 32.56%      Mismatches: 47
Query Match:      7.55%      Indels:      23
DB:              4          Gaps:      5

US-09-854-122-16 (1-391) x US-09-795-926-7 (1-447)
Qy      82 AspAsnPhelysAlaArgGluThrAspIleIleLeuThrThrLeuProLysAlaGlyThr 101
Db      31 GACACCTTCGAAGCCAGACATGATGACATGCTGTAGCATCTTATCCAAAGTCGGTTCA 90
Qy      102 ThrTrp-----ThrLysAlaLeuThrPheAlaIleLeuThrArgAspValAsn 117
Db      91 AACTGGATTCTCCACATGTCAGTGAATTAATATATGCTGTCTTAAATAAGTATAAA 150
Qy      118 HisProSerSerProThrHisProLeuLeuPheAsnProHisSerCysValGlnAsn 137
Db      151 TATCCAGAAATCCAGTT-----CTTGAATGTGGGATTCA 186
Qy      138 LeuGluTyrLeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsnGlu 157
Db      187 GAAAAATAT-----CAGAGAAATGAAAGGCTTTCCA-----216
Qy      158 SerProArgLeuPheAlaGlyHisIleProTyrSerLeuLeuProAlaSerValLeuLys 177
Db      217 TCACCAAGGATTTGGCAACTCACCCTCATATGACAAATACCTGGGTCTATCTTCGAG 276
Qy      178 SerGlyThrLysIleIleAsnIleSerArgAsnArgLysSerThrPheValSerPheTrp 197
Db      277 AATAAGCCAGATATTGGTGATATTTCCGAACCCATAAGATACAGCAGTATCTCTTTTG 336
Qy      198 LysPheGlyAsnLeuIleAsnProAsp

```

```

Db      337 CATTTCCACACGATGTC---CCCGAT 360

RESULT 14
US-09-795-926-11
; Sequence 11, Application US/09795926
; Patent No. 6555669
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kieke, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 561
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-795-926-11

Alignment Scores:
Pred. No.:      3,05e-09      Length:      561
Score:          155.50      Matches:      42
Percent Similarity: 45.74%      Conservative: 17
Best Local Similarity: 32.56%      Mismatches: 47
Query Match:      7.55%      Indels:      23
DB:              4          Gaps:      5

US-09-854-122-16 (1-391) x US-09-795-926-11 (1-561)
Qy      82 AspAsnPhelysAlaArgGluThrAspIleIleLeuThrThrLeuProLysAlaGlyThr 101
Db      145 GACACCTTCGAAGCCAGACATGATGACATGCTGTAGCATCTTATCCAAAGTCGGTTCA 204
Qy      102 ThrTrp-----ThrLysAlaLeuThrPheAlaIleLeuThrArgAspValAsn 117
Db      205 AACTGGATTCTCCACATGTCAGTGAATTAATATATGCTGTCTTAAATAAGTATAAA 264
Qy      118 HisProSerSerProThrHisProLeuLeuPheAsnProHisSerCysValGlnAsn 137
Db      265 TATCCAGAAATCCAGTT-----CTTGAATGTGGGATTCA 300
Qy      138 LeuGluTyrLeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsnGlu 157
Db      301 GAAAAATAT-----CAGAGAAATGAAAGGCTTTCCA-----330
Qy      158 SerProArgLeuPheAlaGlyHisIleProTyrSerLeuLeuProAlaSerValLeuLys 177
Db      331 TCACCAAGGATTTGGCAACTCACCCTCATATGACAAATACCTGGGTCTATCTTCGAG 390
Qy      178 SerGlyThrLysIleIleAsnIleSerArgAsnArgLysSerThrPheValSerPheTrp 197
Db      391 AATAAGCCAGATATTGGTGATATTTCCGAACCCATAAGATACAGCAGTATCTCTTTTG 450

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model  
Run on: March 20, 2004, 03:52:24 ; Search time 444 Seconds  
(without alignments)  
3257.588 Million cell updates/sec

Title: US-09-854-122-16  
Perfect score: 2059  
Sequence: 1 TRGITGIAVACLPLINAGIL.....CLLSYSFVNNFILLKKKK 391

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2438257 seqs, 1849576744 residues

Total number of hits satisfying chosen parameters: 4876514

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model US09854122/runat 18032004 100001 3284/app query.fasta 1.583  
-Q=/cgn2\_1/USPTO.spool US09854122/runat 18032004 100001 3284  
-DB=Published Applications NA -QFMT=fastap -SURFIX=rbnp -MINMATCH=0.1  
-LOPECL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62  
-TRANS=human40.cgi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US09854122 @CGN 1 1 221 @runat 18032004 100001 3284  
-NCPU=6 -ICPU=3 -NO\_WMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DSLOP=6 -DELEX=7

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/PCTUS\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
18: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
-----					

1	1996	96.9	1192	9	US-09-854-122-15	Sequence 15, Appl
2	589	28.6	1311	12	US-10-424-599-58971	Sequence 58971, A
3	545	28.5	996	9	US-09-938-842A-1027	Sequence 1027, Ap
4	545	26.5	996	11	US-09-938-842A-1027	Sequence 1027, Ap
5	537.5	26.1	981	9	US-09-938-842A-2049	Sequence 2049, Ap
6	537.5	26.1	981	11	US-09-938-842A-2049	Sequence 2049, Ap
7	527	25.6	271990	14	US-10-195-144-87	Sequence 87, Appl
8	527	25.6	271990	15	US-10-345-072-87	Sequence 87, Appl
9	525.5	25.5	1403	12	US-10-425-114-21478	Sequence 21478, A
10	525.5	25.5	1526	12	US-10-425-114-27795	Sequence 27795, A
11	521.5	25.3	1017	15	US-10-259-194A-3	Sequence 3, Appli
12	519.5	25.2	1047	15	US-10-259-194A-379	Sequence 379, App
13	501.5	24.4	984	14	US-10-259-165-443	Sequence 443, App
14	501.5	24.4	984	14	US-10-259-165-443	Sequence 443, App
15	483	23.5	2067	14	US-10-195-144-80	Sequence 80, Appl
16	483	23.5	2067	15	US-10-345-072-80	Sequence 80, Appl
17	479	23.3	1032	15	US-10-259-194A-329	Sequence 329, App
18	446.5	21.7	1041	14	US-10-259-165-331	Sequence 331, App
19	444.5	21.6	787	12	US-10-424-599-8242	Sequence 8242, A
20	443.5	21.5	1461	15	US-10-259-194A-131	Sequence 131, App
21	443.5	21.5	1461	15	US-10-260-238-1003	Sequence 1003, Ap
22	424	20.6	1050	15	US-10-259-194A-15	Sequence 15, Appl
23	321	15.6	515	10	US-09-770-961-120	Sequence 120, App
24	268	13.0	1004	13	US-10-199-330-3	Sequence 3, Appli
25	268	13.0	1004	14	US-10-199-334-3	Sequence 3, Appli
26	268	13.0	1004	14	US-10-199-329-3	Sequence 205, App
27	266.5	12.9	1167	12	US-10-072-012-205	Sequence 1, Appli
28	264	12.8	942	13	US-10-199-330-1	Sequence 1, Appli
29	264	12.8	942	14	US-10-199-329-1	Sequence 1, Appli
30	264	12.8	942	14	US-10-199-329-1	Sequence 1, Appli
31	260	12.6	922	13	US-10-199-330-2	Sequence 2, Appli
32	260	12.6	922	14	US-10-199-334-2	Sequence 2, Appli
33	260	12.6	922	14	US-10-199-329-2	Sequence 2, Appli
34	257	12.5	1363	9	US-09-917-800A-1391	Sequence 1391, Ap
35	257	12.5	1363	9	US-10-388-934-204	Sequence 204, App
36	255	12.4	956	15	US-10-116-275-232	Sequence 232, App
37	253.5	12.3	921	9	US-09-898-570-25	Sequence 25, Appl
38	253.5	12.3	921	10	US-09-839-446-25	Sequence 25, Appl
39	253	12.3	462	15	US-10-259-194A-643	Sequence 643, App
40	253	12.3	462	15	US-10-259-194A-492	Sequence 492, App
41	250	12.1	1000	15	US-10-388-934-493	Sequence 493, App
42	250	12.1	1000	15	US-10-388-934-789	Sequence 789, App
43	250	12.1	1000	15	US-10-388-934-825	Sequence 825, App
44	248.5	12.1	1229	9	US-09-816-790-19	Sequence 19, Appl
45	245.5	11.9	1138	9	US-09-981-353-100	Sequence 100, App

#### ALIGNMENTS

#### RESULT 1

US-09-854-122-15  
Sequence 15, Application US/09854122  
Patent No. US20020016980A1  
GENERAL INFORMATION:  
APPLICANT: SMITH, ROBERT  
TITLE OF INVENTION: TRANSGENIC PLANTS INCORPORATING TRAIT OF ZOSTERA MARINA  
FILE REFERENCE: PHA-007.01  
CURRENT APPLICATION NUMBER: US/09/854,122  
PRIOR FILING DATE: 2001-09-10  
PRIOR FILING DATE: 2000-05-10  
NUMBER OF SEQ IDS: 51  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 15  
LENGTH: 1192  
TYPE: DNA  
ORGANISM: Zostera marina  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1..33, 37..1041, 1045..1059, 1063..1077,  
LOCATION: 1081..1086, 1090..1119, 1123..1191)  
US-09-854-122-15

Alignment Scores: 2.19e-230 Length: 1192  
Pred. No.: 1996.00 Matches: 391  
Score: 1996.00 Matches: 391  
Percent Similarity: 98.49% Conservative: 0  
Best Local Similarity: 98.49% Mismatches: 0  
Query Match: 96.94% Indels: 6  
DB: 9 Gaps: 6

US-09-854-122-16 (1-391) x US-09-854-122-15 (1-1192)

QY 1 ThrArgGlyLeuThrGlyLeuAlaValAlaCys---LeuProLeuIleMetAlaGlyLeu 19  
DB 1 ACCGGGGAATACCTGGATGGCTGCTGCTGCTAGCTACCTACCTGATGATGGCTGGAT 60  
QY 20 LeuAlaLeuGluLysCysPheGlySerLysAsnGlnGlnLysGluGluAspSerLys 39  
DB 61 TTAGCTTTGGAGAAATGTTCCGATCCAGAAATGAGCAAGAGAGGAAGAGATTCACAA 120  
QY 40 MetTyrLysArgTyrArgGluIleValSerSerLeuProSerAsnAspTyrTrpGlyAsp 59  
DB 121 ATGTACAGAGATATAGAGATGTTCTTCCTCCTCCCTGATGATGATGATGGGGAT 180  
QY 60 ThrMetArgLeuTyrLysGlyPheTrpGlnMetGlyTyrLeuValProGlyIleMetAla 79  
DB 181 ACCATGAGGTTGTACAGGGATTTTGGCAATGGGATATCTTGTACCTGGTATCATGGCT 240  
QY 80 PheGluAspAsnPheIleAlaArgGluThrAspIleLeuThrLeuProLysAla 99  
DB 241 TTCGAAGATAATTTCAAGCTCGAGACGGACATATTCCTTACGACTCTTCCAAAGGCT 300  
QY 100 GlyThrThrTrpThrLysAlaLeuThrPheAlaIleLeuThrArgAspValAsnHisPro 119  
DB 301 GGACGACATGACGAGGACCTGCTGCTTGGCCCTTAACACGAGATGTTAACCCACA 360  
QY 120 SerSerProThrHisProLeuLeuPheAsnProHisSerCysValGlnAsnLeuGlu 139  
DB 361 TCATCCACGACATCCATCTTTGTTCTTCAACCTCATCTGTTGTTTCAAAATTTGGAG 420  
QY 140 TyrLeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsnGluSerPro 159  
DB 421 TATTTGATCATGGTAGAGAAATACGATGCCAGACCTCGATATGTTGAATGAATCGCGC 480  
QY 160 ArgLeuPheAlaGlyHisIleProTyrSerLeuLeuProAlaSerValLeuLysSerGly 179  
DB 481 AGGTTGTTGCCGACACATCCCATCTCTTGTGCGCGCTGCTTTTGAATCGGA 540  
QY 180 ThrLysIleIleAsnIleSerArgAsnArgLysSerThrPheValSerPheTrpLysPhe 199  
DB 541 ACAAAATCATCATATAGCCGCAACCGTAAGATACATTGTTGCTTTTGGAAATTT 600  
QY 200 GlyAsnLeuIleAsnProAspLysLeuLeuAspLeuGluLysSerValAspIlePheAla 219  
DB 601 GGCAATCTGATTAACCCGCAAGTATTATGGACCTCGAAAGAGCGTTGATATCTTCGA 660  
QY 220 SerGlyLysSerPheCysGlyProGluTrpAsnPheGlnAlaGluPheThrAsnAlaLa 239  
DB 661 TCGGAATCTCTTTTGTGACCGGATGGAATTTTCAAGCGGAGTTTCAATCGGGG 720  
QY 240 SerThrAsnSerAsnLeuLeuLeuSerTyrGluGluMetLeuGluLysProValGlu 259  
DB 721 TCTACTAATTCAAACTTGCTATTGTTGAGTTAGCAAGAAATGTTAGAGAGCGAGTTGAA 780  
QY 260 AsnValLysLysLeuAlaGluPheMetGlyCysGlyPheThrAspAspGluGluLysGln 279  
DB 781 AATGTGAAGAGCTAGCTGATTCATCGGATGTTGGGTTTACACCGATGAGGAGAACAA 840  
QY 280 GlyIleValAspGluIleValLysLeuCysSerPheAspAsnLeuLysAsnGlnVal 299  
DB 841 GGGATTGTTGATGAGATAGTAAACTTTGAGCTTCGACATCTGAAAGATCAACAGGTG 900  
QY 300 AsnLysAsnGlySerSerTyrAsnSerLysIleAspAsnLysHisPhePheArgLysGly 319

RESULT 2

US-10-424-599-58871  
; Sequence 58871, Application US/10424599  
; Publication No. US20040031072A1

GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 58871  
; LENGTH: 1311

TYPE: DNA

; ORGANISM: Glycine max

FEATURE:

; NAME/KEY: unsure

; LOCATION: (1)...(1311)

; OTHER INFORMATION: unsure at all n locations

FEATURE:

; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_24172C.1

US-10-424-599-58871

Alignment Scores:

Fred. No.: 1.95e-60 Length: 1311  
Score: 589.00 Matches: 125  
Percent Similarity: 59.88% Conservative: 69  
Best Local Similarity: 38.58% Mismatches: 116  
Query Match: 28.61% Indels: 14  
DB: 12 Gaps: 8

US-09-854-122-16 (1-391) x US-10-424-599-58871 (1-1311)

QY 25 CysPheGlySerLysAsnGluGlnGluLysGlu-----GluAspSerLys 39  
DB 60 TGCTTCAGAGAAGAAATGAATCCGAGAAAGGGAGAAATAACAATAGAGAGACAG 119  
QY 40 MetTyrLysArgTyrArgGluIleValSerSerLeuProSerAsnAspTyrTrpGlyAsp 59  
DB 120 CTAAGTCAAGATGATAGGAGTTGATACTCTCTCTCTAGGAGAGAGAGTTGGAGAACA 179  
QY 60 Thr---MetArgLeuTyrLysGlyPheTrpGlnMetGlyTyrLeuValProGlyIleMet 78  
DB 180 CGTTATATATATATATTTCAAGGATTTTGGTGCAGCCATTTGGAATCCAGCAATATC 239  
QY 79 AlaPheGluAspAsnPheLysAlaArgGluThrAspIleIleLeuThrThrLeuProLys 98  
DB 240 ACTTTTCAGAGCACTTCCAAAGCTTAAGACAGTGTATGTTGGGCCACCAATCCAAA 299  
QY 99 AlaGlyThrThrTrpThrLysAlaLeuThrPheAlaIleLeuThrArgAspValAsnHis 118



```
QY 331 GluMetIleLysLysLeuGlu 337
Db 928 CAAATGGCGAAACCTTTGAT 948

RESULT 4
US-09-938-842A-1027
; Sequence 1027, Application US/09938842A
; Publication No. US20040009476A9
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SAME, AND METHODS OF USE
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 1027
; LENGTH: 996
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-1027

Alignment Scores:
Pred. No.: 2.51e-55 Length: 996
Score: 545.00 Matches: 118
Percent Similarity: 58.31% Conservativity: 61
Best Local Similarity: 26.44% Mismatches: 118
Query Match: 38.47% Indels: 10
DB: 11 Gaps: 8

US-09-854-122-16 (1-391) x US-09-938-842A-1027 (1-996)

QY 36 GluAspSerLysMetTyrLysArgTyrArgGluIleValSerLeuProSerAsn--- 54
Db 43 AAAGACGACACAGCTTAGTACGAAGAACCAAGCTTGTATCACTTCTCTACCTTCAGACAA 102
QY 55 AspTyrTrpGlyAspThrMetArgLeuTyrLysGlyPheTyrGlnMetGlyTyrLeuVal 74
Db 103 GATTTCATGGGTATGGTCTCTACAACTACAAAGTTGTTGGTACTATCCAAACACACTC 162
QY 75 ProGlyIleMetAlaPheGluAspAsnPhenylsAlaArgGluThrAspIleLeuThr 94
Db 163 CAAGCGTTCTTGAGCTCCAAACACACTTCAAGCCACGAGATCTGATATATATCTCGCT 222
QY 95 ThrLeuProLysAlaGlyThrThrThrThrLysAlaLeuThrPheAlaIleLeuThrArg 114
Db 223 TCTTTGCCAAAGGTGGAACCACTTGGCTCAATCCCTAAATTTTCGCTGTTGATACAGA 282
QY 115 AspValAsnHisProSerSerProThrHisProLeuLeuPheAsnProHisSerCys 134
Db 283 GAAAGTACCGCGGAACCCCTCAACACATCTTTGCTCTTACAAACCCCTCATGACCTT 342
QY 135 ValGlnAsnLeuGluTyr---LeuTyrMetGlyArgGluAsnThrMetProAspLeuAsp 153
Db 343 GTCCCATTTCTTGGAGTTGAGTTATACGCTAATAGCCAA-----ATTCCGGATCTCGCA 396
QY 154 MetLeuAsnGluSerProArgLeuPheAlaGlyHisIleProTyrSerLeuLeuProAla 173
Db 397 AGATATCT---TCTCTATGATCTTTCTACACATGCTACACATGCTTACGAGTGGTGA 453
QY 174 SerValLeuLysSerGlyThrLysIleLeuAsnIleSerArgAsnArgLysSerThrPhe 193
Db 454 GCCACCACAAAGCT---TGCAAAACCGTATATGTGTGTAGAGGTATCAAGATACGTTT 510
```

```
QY 194 ValSerPheTrpLysPheGlyAsnLeuIleAsnProAspLysLeu-----LeuAspLeu 211
Db 511 GTCTCGGCTGGCATTATAGAAACATGTTGCATCGCACCAAGATGGATCAAGCCACTTTT 570
QY 212 GluLysSerValAspIlePheAlaSerGlyIleSerPheCysGlyProGluTrpAsnPhe 231
Db 571 GAGCTCATGTTTGTATGCTTATTTAGAGGAGTCTCTTATATGACCTTATTGGGAACAT 630
QY 232 GlnAlaGluPhe---ThrAsnAlaAlaSerThrAsnSerAsnLeuLeuLeuSerTyr 250
Db 631 GTATTGAGCTATTGGAAAGGGAGCTTGGAAAGCAAGGAGATGTTCTTTTCATGAAGTAC 690
QY 251 GluGluMetLeuGluLysProValGluAsnValLysLysLeuAlaGluPheMetGlyCys 270
Db 691 GAAGAGATTAATTGAGAGCCCTCGTGTCAAGTCAAGAGACTCGCCGAGTCTTGGATGT 750
QY 271 GlyPheThrAspGluGluLysGlnGlyIleValAspGluIleValLysLeuCysSer 290
Db 751 CCATTCCACGAAGAAAGAAAGTGGATCGGTGGAGAGATCTTGAAGTTGTGTAGT 810
QY 291 PheAspAsnLeuLysAsnGlnValAsnLysAsnGlySerSerTyrAsnSerLysIle 310
Db 811 TTACGAAATTTAAGCAATTTGGAGGTTATATAGATGGGACACG---AGAATTGGTGT 867
QY 311 AspAsnLysHisPhePheArgLysGlyValArgAspTrpAlaAsnTyrLeuThrSer 330
Db 868 GATTCTCAGGTCTTCTTTAGGAAGGTGAAGTTGGTGTATGGAAGATCATCTTACGCCA 927
QY 331 GluMetIleLysLysLeuGlu 337
Db 928 CAAATGGCGAAACCTTTGAT 948
```

```
RESULT 5
US-09-938-842A-2049
; Sequence 2049, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SAME, AND METHODS OF USE
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 2049
; LENGTH: 981
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-2049
```

```
Alignment Scores:
Pred. No.: 1.97e-54 Length: 981
Score: 537.50 Matches: 123
Percent Similarity: 57.10% Conservativity: 54
Best Local Similarity: 39.68% Mismatches: 112
Query Match: 26.10% Indels: 21
DB: 9 Gaps: 9

US-09-854-122-16 (1-391) x US-09-938-842A-2049 (1-981)

QY 37 AspSerLysMetTyrLysArgTyrArgGluIleValSerSerLeuProSerAsnAspTyr 56
Db 37 GATGAAGATCTGACACAGAAACAGAGCTCTGATCTTCTTCTCTTAAGAAAGAAAGGT 96
```







```
; PRIOR APPLICATION NUMBER: 60/305,363
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: 60/308,736
; PRIOR FILING DATE: 2001-07-30
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 87
; LENGTH: 271990
; TYPE: DNA
; ORGANISM: Raphanus sativum
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (144241)..(144300)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-345-072-87

Alignment Scores:
Pred. No.: 5,34e-49 Length: 271990
Score: 527.00 Matches: 120
Percent Similarity: 57.81% Conservative: 65
Best Local Similarity: 37.50% Mismatches: 115
Query Match: 25.59% Indels: 21
DB: 15 Gaps: 8

US-09-854-122-16 (1-391) x US-10-345-072-87 (1-271990)

QY 29 LysAsnGlnGluGlyLysGluGlu-----AspSerLys 39
Db 244900 AAGAACCAAGAACACATGATCGGAGGAGCTTCGTCGACCTTAGACGACGACAG 244841
QY 40 MetTyrLysArgTyrArgGluLeuValSerSerLeuProSerAsn---AspTyrTrpGly 58
Db 244840 ATAAGTGAAGAAACACGAGGAGGTGATCTCTTCGCTTCCTTCACACACGATACCAAGG 244781
QY 59 AspThrMetArgLeuTyrLysGlyPheTrpGlnMetGlyTyrLeuValProGlyLeuMet 78
Db 244780 CATPAGATTGTAAATCATCAAGGATGTGTGTATATTATACACACCTCCCAAGGTGCTCA 244721
QY 79 AlaPheGluAspAsnPhenylsAlaArgGluThrAspIleLeuThrTrpLeuProLys 98
Db 244720 AT-TTCAGAGAGGTTTCAACCGCAAGACACTGATGATATCATTCGTCGACCCCAA 244662
QY 99 AlaGlyThrTrpThrLysAlaLeuThrPheAlaIleLeuThrArgAspValAsnHis 118
Db 244661 TCAGGCACACTTGGCTCAAGGCCCTCACAGTCGCTCTGTCGAGAGATCAAGAACCCAC 244602
QY 119 ProSerSerProThrHisProLeuLeuPhePheAsnProHisSerCysValGlnAsnLeu 138
Db 244601 TCTTCTGAT-----CATCCTCTCTATATCATATCTTCATGCGCATTTATACCATTC 244548
QY 139 GluTyrLeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsnGluSer 158
Db 244547 GAG---ATCGATGTGTACCAAGAAAGCTCAAGTCTCAACCTAGCCAAAGTTCTCAGCACCT 244491
QY 159 ProArgLeuPheAlaGlyHisIleProTyrSerLeuLeuProAlaSerValLeuLysSer 178
Db 244490 CCAGAGCTGTCTCGACTCATCTCCCTCCCTCACAGCATCCACCAAGCATCAGACACTCT 244431
QY 179 GlyThrLysIleLeuAsnIleSerArgAsnArgLysSerThrPheValSerPheTrpLys 198
Db 244430 CCTTGCAGATTGTGTACGTGTGCAGGAACGTGAAGGACACGTTGATCTCGTGTGGTTT 244371
QY 199 PheGly-----AsnLeuIleAsnProAspLysLeuLeuAspLeuLysSerVal 215
Db 244370 TACAGCTGTCTATATATAAATCGAACCAACCAAGAGAGATT---CTCAGATCTATGTTT 244314
QY 216 AspIlePheAlaSerGlyIleSerPheCysGlyProGluTrpAsnPheGlnAlaGluPhe 235
Db 244313 AACGAGTCTCGCATGCAACCAACTATTTGGACCTTTTGGATCATCTCTTGTAGTTAC 244254
QY 236 ThrAsnAlaIleSerThrAsn---SerAsnLeuLeuLeuSerTyrGlnGluMetLeu 254
Db 244253 TGGAGAGGAGCTTGGGAAGACCCCAAGCATGTCTCTTTCATGAGTATGAGGAGATGAAA 244194
```

```
QY 255 GluLysProValGluAsnValLysLysLeuAlaGluPheMetGlyCysGlyPheThrAsp 274
Db 244193 GCCGAGCCTCGTGTATCATCAAGAGACTTGGGACTTCTTGGGATGCTCCTTTACTAAG 244134
QY 275 AspGluGluLysGlnGlyIleValAspGluLeuValLysLeuCysSerPheAspAsnLeu 294
Db 244133 CAAGAAGAAGATAGTGTATCTGTGGACGGGATCTTGGACTCTGCTCTCTGCGTAATCTG 244074
QY 295 LysAsnGlnValAsnLysAsnGlySerSerTyrAsnSerLysIleAspAsnLysHis 314
Db 244073 AGCAGTTTGGAGGTAACAAACAGGACAAATAACAAT-----GTGGAGCACAGTTT 244020
QY 315 PhePheArgLysGlyGluValArgAspTrpAlaAsnTyrLeuThrSerGluMetIleLys 334
Db 244019 TTTTTCGTAAGGAGAAGTCGGTGACTCGAAAAATTTATCTTACGCTCGAATGGAGAAA 243960
```

```
RESULT 9
US-10-425-114-21478
; Sequence 21478, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 21478
; LENGTH: 1403
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3353-009-H8_FLI
US-10-425-114-21478
```

```
Alignment Scores:
Pred. No.: 1.02e-52 Length: 1403
Score: 525.50 Matches: 129
Percent Similarity: 50.49% Conservative: 77
Best Local Similarity: 31.62% Mismatches: 134
Query Match: 25.52% Indels: 69
DB: 12 Gaps: 15
```

```
US-09-854-122-16 (1-391) x US-10-425-114-21478 (1-1403)
QY 30 AsnGluGlnGluLysGluGluAsp-----SerLysMetTyrLysArgTyrArgGluLeu 47
Db 148 AATGCACAAGAAGCAGGATCACCATTGGCGAGTCTCTCTCACTCCACATACATAGCCAGCATC 207
QY 48 ValSerSerLeuProSerAsnAspTyrTrpGlyAspThr-----Met 61
Db 208 ATCCCTTCGCTGCCCTCTC-----GAGACAGGTGCCACCGTTCCTCTC 252
QY 62 ArgLeuTyrLysGlyPheTrpGlnMetGlyTyrLeu-----ValProGlyLeu 77
Db 253 CGCGGTACGCCAACTTTTGGTGTCTGAGGTAATATTGAAGGCAGACCTCCCGGCATT 312
QY 78 MetAlaPheGluAspAsnPhenylsAlaArgGluThrAspIleLeuThrThrLeuPro 97
Db 313 -----CANTCTGCTTCAAGCCGAGACCCAGCGGTTCGCGAAGCTTCCCC 363
QY 98 LysAlaGlyThrThrTrpThrLysAlaLeuThrPheAlaIleLeuThrArgAspValAsn 117
Db 364 AAGTCTGCACCACTTGGCTCAAGCCCTTGGCTTCGCGAGCGTGAAGCGTTCACCGCAT 423
QY 118 HisProSerSerProThrHisProLeuLeuPhePheAsnProHisSerCysValGlnAsn 137
```

```
Db 424 CCGCCATTGGATGGCGACCATCCGCTCGTGTGCAACCCCTCATGACTGTGTGACGGTTC 483
Qy 138 LeuGluTyrLeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsnGlu 157
Db 484 CTCGATGCTAATTTCAACCCAGCAGAGATGAGTGGAGGCTCTC-----CCG 531
Qy 158 SerProArgLeuPheAlaGlyHisIleProTyrSerLeuLeuProAlaSerValLeu--- 176
Db 532 TCCCGCGCGGTGTAGCCACGACGCTCCCTCTACTCCCTGCTGCGCGCAGCATCACCGGG 591
Qy 177 -----LysSerGlyThrLysIleLeuAsnIleSerArgAsnArgLysSerThrPhe 193
Db 592 GATAGGAGGCGCTCGGGTGGCGATGCTGTATGCTGCGCGGACCCAGGAGCGCGCTG 651
Qy 194 ValSerPheThrLysPhe-----GlyAsnLeuLeuAsnProAsp---LysLeu 208
Db 652 GTCTCTGCTGCTGCTTTCACGAGGAAGCGCGCTCGCGCTGGGGGTGACGCTCGATCG 711
Qy 209 LeuAspLeuGluLysSerValAspIlePheAlaSerGlyIleSerPheCysGlyProGlu 228
Db 712 TTCACGATCCAGGAGCGCTTGGAGCTTTCTGATGCTGCTGATGCTGCGCGCGCGCAG 771
Qy 229 TrpAsnPheGlnAlaGluPhe---ThrAsnAlaAlaSerThrAsnSerAsnLeuLeu 247
Db 772 TGGGAGCAGCTCCTCCAGTACTGGGAGGAGGAGCGCTGAGGAGCGCTGACAGGGTGTTC 831
Qy 248 LeuSerTyrGluGluMetLeuLysProValGluAsnValLysLysLeuAlaGluPhe 267
Db 832 CTCGATGACGAGGAGATGCTGATTGATCCCGAGGACACGTCAGAGAGCTCGCCAAATTC 891
Qy 268 MetGlyCysGlyPheThrAspAspGluLysGlnGlyIleValAspGluIleValLys 287
Db 892 ATGGGGTGTGATTTTCTGAGGAGGAGGAGGAGCAGCGGGTGTGTGAGCGGCATCGTGGAG 951
Qy 288 LeuCysSerPheAspAsnLeuLysAsnGlnValAsnLysAsnGlySerSerTyrAsn 307
Db 952 CTGTGTACCTCGGCAAGATGAGGACATGAGGAGTGAACAGAAATGGAGCAATATGTTG 1011
Qy 308 SerLysIleAspAsnLysHisPhePheArgLysGlyGluValArgAspTrpAlaAsnTyr 327
Db 1012 GGG---GTCAGAACGAAAGCTACTTTCAGGAGGAGGAGTGTGCGGACTGGAGCAACCAT 1068
Qy 328 LeuThrSerGluMetIleLysLysLeuGlu-----Thr 338
Db 1069 ATGACGCGGACATCGCGCAGAGGCTAGACAAAGTGTGATGAGGATGCAATGCAAGGACT 1128
Qy 339 AlaGlyLysIleAsnGluSer-----GluLys 347
Db 1129 GGATT-CAGCTTACAGCAAGCATGATGACCACTCGCTCATGTTGAGTGGAACTC 1187
Qy 348 HisLeuLeuSerAsnLysAsnLeuThrAsnPheAsnLeu----- 360
Db 1188 CATAATGCCAAGAACAACTCGTTTGTGCGTATGCTGCTGGATGGAGATGATACCAAGC 1247
Qy 361 -----AsnAsnTyrCysGluLysSerAsnIleSerLeuLysLeuSer 375
Db 1248 CAGCCATGTGAGCATCTCTTTTGTGATGAGGGAATGATTTCGCTATATTCAT 1307
Qy 376 ---TyrSerPheValIleAsnAsn 382
Db 1308 CTGTATAACTTGTACAAACCAT 1331

RESULT 10
US-10-425-114-27795
; Sequence 27795, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
```

```
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425.114
; NUMBER OF SEQ ID NOS: 2003-04-28
; SEQ ID NO 27795
; LENGTH: 1526
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: Lib4740-009-D3_FLI
US-10-425-114-27795
```

```
Alignment Scores: 1,178-52 Length: 1526
Pred. No.: 525.50 Matches: 129
Score: 50.49% Conservative: 77
Percent Similarity: 31.62% Mismatches: 134
Best Local Similarity: 25.52% Indels: 69
Query Match: 12 Gaps: 15
DB:
```

US-09-854-122-16 (1-391) x US-10-425-114-27795 (1-1526)

```
Qy 30 AsnGluGlnGluLysGluLysAsp-----SerLysMetTyrLysArgTyrArgGluIle 47
Db 231 AATGCACAAAGACGACGATCACCATTGGCGAGTCTCTCTCACTCCAAACATAGCCAGCATC 290
Qy 48 ValSerSerLeuProSerAsnAspTyrTrpGlyAspThr-----Met 61
Db 291 ATCCCTTCCTCCCTCCTC-----GAGACGAGGTGCCACCGTTCCTCTCCTC 335
Qy 62 ArgLeuTyrLysGlyPheTrpGlnMetGlyTyrLeu-----ValProGlyIle 77
Db 336 CGCGCGTACGCAACTTTTGGTGGCTGAGGTATATTGAAGGCGAGACCTCCCGCGCAT 395
Qy 78 MetAlaPheGluAspAsnPheLysAlaArgGluThrAspIleLeuThrThrLeuPro 97
Db 396 -----CATTCCTGCTTCAAGCGGAGACCCACCGAGCTTGTGTCGCAAGCTTCC 446
Qy 98 LysAlaGlyThrThrTrpThrLysAlaLeuThrPheAlaIleLeuThrArgAspValLeu 117
Db 447 AAGTCTGGCACCATTTGGCTCAAGGCCCTTGGCTTGGCGACGCTGAAGCGTTCACG 506
Qy 118 HisProSerSerProThrHisProLeuLeuPhePheAsnProHisSerCysValGlnAsn 137
Db 507 CCGCCATTGGATGGCGACCATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 566
Qy 138 LeuGluTyrLeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsnGlu 157
Db 567 CTCGATGCTAATTTCAACCCAGCAGAGGATGAGTTGGAGGCTCTC-----CCG 614
Qy 158 SerProArgLeuPheAlaGlyHisIleProTyrSerLeuLeuProAlaSerValLeu--- 176
Db 615 TCCCCGCGGTGTAGCAGCAGCACCTCCCTCTACTCCCTGCTGCTGCTGCTGCTGCTGCTGCT 674
Qy 177 -----LysSerGlyThrLysIleLeuAsnIleSerArgAsnArgLysSerThrPhe 193
Db 675 GATAGGAGGCGCTCGGGTGGCGATGCTGTGCTGCGCGGAAACCCAGGAGCGCGCTG 734
Qy 194 ValSerPheThrLysPhe-----GlyAsnLeuLeuAsnProAsp---LysLeu 208
Db 735 GTCTCTGCTGCTGCTTTCACGAGGAAGCGCGCTCGCGCTGCGCGCTGCTGCTGCTGCTGCT 794
Qy 209 LeuAspLeuGluLysSerValAspIlePheAlaSerGlyIleSerPheCysGlyProGlu 228
Db 795 TTCACGATCCAGGAGCGCTTGGAGCTTTTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 854
Qy 229 TrpAsnPheGlnAlaGluPhe---ThrAsnAlaAlaSerThrAsnSerAsnLeuLeu 247
Db 855 TGGGAGCAGCTCTCCAGTACTGGGAGGAGGAGCGCTGAGGAGCGCTGACAGGGTGTTC 914
```

```
QY 248 LeuSerTyrGluGluMetLeuGluLysProValGluAenValLysLysLeuAlaGluPhe 267
DB 915 CTTGGTACGAGAGATGCTGATTGATCCCGAGGACACAGCTCAGGAAGTCCCAAGTTC 974
QY 268 MetGlyCysGlyPheThrAspGluGluLysGlnGlyLeuValAspGluLysValLys 287
DB 975 ATGGGTGTGGATTCTCAGGAGGAGGAGGACGCGGGTGTGAGCGCCATCGTGGAG 1034
QY 288 LeuCysSerPheAspAsnLysLysAenGlnGlnValAsnLysAenGlySerSerTyrAsn 307
DB 1035 CTGTGTAGCTGGGCAAGATGAGGACATCGAGGTGAACAGAAATGGAAGCAATATGTTG 1094
QY 308 SerLysLeuAspAsnLysHisPhePheArgLysGlyGluValArgAspTyrAlaAsnTyr 327
DB 1095 GGG---GTCAGAACCAAGAGCTACTTCAGAGAGGAGTTGCTGGGACATGGAGCAACAT 1151
QY 328 LeuThrSerGluMetLysLysLysLeuGlu-----Thr 338
DB 1152 ATGACCGGACATGCGCAGAGGCTAGACAAGGTGCTAGAGGATGCAATTGCAAGGGACT 1211
QY 339 AlaGlyLysLeuAsnGluSer-----GluLys 347
DB 1212 GATT-CAGCTTTACGACCAAGCATGATGACCACTGCGCTCATGTTAGTTGCGAATC 1270
QY 348 HisLeuLeuSerAsnLysAsnLeuThrAsnPheAsnLeu----- 360
DB 1271 CATAATGCCAAGAACCACTGCTTTGCTGCTATTGCTTGTGATGGAGATCATACCAGC 1330
QY 361 -----AsnAsnTyrCysGluLysSerAsnLysSerAsnLysSerLysLeuLeuSer 375
DB 1331 CAGCCATGGTGAGCATCTCTCTTTTGTGATGAAGGGGAAATTGATTGCTGTAATTCAT 1390
QY 376 ---TyrSerPheValLeuAsnAsn 382
DB 1391 CTGTATAACTTTGTTACAAACCAT 1414
```

## RESULT 11

```
US-10-259-194A-3
; Sequence 3, Application US/10259194A
; Publication No. US20040010815A1
; GENERAL INFORMATION:
; APPLICANT: Lange, Markus B.
; APPLICANT: Ghasseman, Majid
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Krepis, Joel
; APPLICANT: Moughamer, Todd
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF PLANT GENES
; FILE REFERENCE: 70029-NP
; CURRENT APPLICATION NUMBER: US/10/259,194A
; CURRENT FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,743
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 662
; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta
; SEQ ID NO 3
; LENGTH: 1017
; TYPE: DNA
; ORGANISM: Oryza sativa
US-10-259-194A-3
```

Alignment Scores:

Pred. No.: 1.79e-52 Length: 1017

```
Score: 521.50 Matches: 118
Percent Similarity: 56.03% Conservative: 54
Best Local Similarity: 38.44% Mismatches: 116
Query Match: 25.33% Indels: 19
DB: 15 Gaps: 8
US-09-854-122-16 (1-391) x US-10-259-194A-3 (1-1017)
QY 42 LysArgTyrArgGluLeuValSerSerLeuProSerAsnAspTyrTyrPheGlyAspThrMet 61
DB 97 GAGGAGTATCCCGCGCGTGTCCAGCCTCCCGAGC-----TATCCCAAGCTGCGCCTG 150
QY 62 ArgLeuTyrLysGlyPheThrPheGlnMetGlyTyrLeuValProGlyLysLeuAlaPheGlu 81
DB 151 CGGCACATACAGGCGCATGTGGCTGATGAGTACACCCCTCCAGGAGTATGAGCATCCAG 210
QY 82 AspAsnPhe---LysAlaArgGluThrAspLysLeuLeuLeuThrLeuProLysLysAlaGly 100
DB 211 CGGAGCTTCGTGCGCGCCCGCCATGGCGAGCTCGTCTCCGAGCGCCCGGTAGTGGCGC 270
QY 101 ThrThrTyrPheLysAlaLeuThrPheAlaLeuThrArgAspValAsnHisProSer 120
DB 271 ACCACTTGGCTCAAGGCGCTGCGCTTCGCGCTTGGCGCGCGCGGCTACTCCCTGCC 330
QY 121 SerProThrHisProLeuLeuPhePheAsnProHisSerCysValGlnAsnLeuGluTyr 140
DB 331 AGTGACCGGATCCGCTCCTCCGCTTTAACCCGACGACTGCGTCCGTTTCATGAGGGG 390
QY 141 Leu-----TyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsn 155
DB 391 CGCATCTCTGAGGATGGGTGTAGATTGATGAGTGGCG----- 432
QY 157 GluSerProArgLeuPheAlaGlyHisLeuProTyrSerLeuLeuProAlaSerValLeu 176
DB 433 ---TCGCGAGGCTCATGTCCACCCACATGCAGCAGCTCTCTACCCAAAGTCCATGCA 489
QY 177 ---LysSerGlyThrLysLeuLeuAsnLysSerArgAsnArgLysSerThrPheValSer 195
DB 490 GATGAACCTGGCTCAAGGTGTATTATTTCAGGAGGCGGAGGAGGATATCTGCTCTCA 549
QY 196 PheTyrLysPheGlyAsnLeuLeuAsnProAspLysLeuLeuAspLeuLysSerVal 215
DB 550 GCCTGGCATCTCTCAGAAATATAGAGCTGAT-----CTTTTCATTTCAAGAAGTGTTC 603
QY 216 AspLeuPheAlaSerGlyLysSerPheCysGlyProGluThrAsnPheGlnAlaGluPhe 235
DB 604 GAGGCTGCTCGGATGGCAAGTCTTTCACCGCGCTATCTGGGATCACATATTGGCTAC 663
QY 236 ThrAsnAlaAlaSerThrAsn---SerAsnLeuLeuLeuSerTyrGluMetLeu 254
DB 664 TGGATGCTTCAAGGCGGACCCAGAGAGGTCTCTGTTCTGTGTATGAGGACCTTCTG 723
QY 255 GluLysProValGluAsnValLysLysLeuAlaGluPheMetGlyCysGlyPheThrAsp 274
DB 724 CGAGATCCGCGCAACATCGTCCAGAACTCCCGACTTCTTGGCCAGCCGCTTCTCATCG 783
QY 275 AspGluGluLysGlnGlyLeuValAspGluLeuValLysLeuCysSerPheAspAsnLeu 294
DB 784 ACCGAGGAGAGAGCTGCGACTGTCAGATATTGAGGCTGTGCAGCTTCGAGAACCTG 843
QY 295 LysAsnGlnGlnValAsnLysAsnGlySerSerTyrAsnSerLysLysLeuAspAsnLysHis 314
DB 844 AAGAGCTTGGAGGTGAATAAGATGGGAGAGCA---TCGTTTCGCTTCCCGAAATCGGTCA 900
QY 315 PhePheArgLysGlyGluValArgAspTyrAlaAsnTyrLeuThrSerGluMetLeuLys 334
DB 901 TACTTCCGAAAGGAGGCGGAGACTGGAAATCCACATGACCGCGGAAATGTTGGAG 960
QY 335 LysLeuGluThrAlaGlyLys 341
DB 961 TGTTTTGACCATCGTCAAG 981
```

RESULT 12

```
US-10-259-194A-379
; Sequence 379, Application US/10259194A
; Publication No. US20040010815A1
; GENERAL INFORMATION:
; APPLICANT: Lange, Markus B.
; APPLICANT: Ghassemlan, Majid
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Moughamer, Todd
; APPLICANT: Provart, Nicholas
; APPLICANT: Rickes, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF PLANT GENES
; FILE REFERENCE: 70029-NP
; CURRENT APPLICATION NUMBER: US/10/259,194A
; PRIOR FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,743
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 662
; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta
; SEQ ID NO 379
; LENGTH: 1047
; TYPE: DNA
; ORGANISM: Oryza sativa
US-10-259-194A-379

Alignment Scores:
Pred. No.: 3.27e-52 Length: 1047
Score: 519.50 Matches: 120
Percent Similarity: 56.38% Conservative: 48
Best Local Similarity: 40.27% Mismatches: 117
Query Match: 25.23% Indels: 13
DB: 8 Gaps: 5

US-09-854-122-16 (1-391) x US-10-259-194A-379 (1-1047)
QY 48 ValSerSerLeuProSerAsnAspTrpTrpGlyAspThrMetArgLeuTyrLysGlyPhe 67
Db 130 GTCCTCGCCCTGCGGGCGGTCTCTCTACGCGCCAGCGGATGGGTCTACGGGGGACC 189
QY 65 TrpGlnMetGlyTyrLeuValProGlyIleMetAlaPhe---GluAspAsnPhelysAla 86
Db 190 TGGGTGTTCGAGAGTGGGCGGCGGCGGTGTGTGCGCATGCGCGGGCGGCTGTGCGCC 249
QY 87 ArgGluThrAspIlelleLeuThrThrLeuProLysAlaGlyThrThrTrpThrLysAla 106
Db 250 CGCGCGGGCGAGTCTCTCTCCAGCTCCCAAGTGGGAGACGCTGCTCAAGGCA 309
QY 107 LeuThrPheAlaIleLeuThrArgAspVal---AsnHisProSerSerProThrHisPro 125
Db 310 CTCGCGTTTCGACCATCGCGCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 369
QY 126 LeuLeuPhePheAsnProHisSerCysValGlnAsnLeuGluTyrLys---MetGly 144
Db 370 CTCGCGCGCTAAACCGGACGAGTGTGCGCTCTCTGACAGGCTCTTCGCGCGCGG 429
QY 145 ArgGluAsnThrMetProAspLeuAspMetLeuAsnGluSerProArgLeuPheAlaGly 164
Db 430 CGCGAGCGGCTTCTCGACGAGTGTG-----CCGTGCGCGGAGGCTCATGTGCGACG 477
QY 165 HisIleProTyrSerLeuLeuProAlaSerValLeu-----LysSerGlyThrLysIle 182
Db 478 CACATCGCGCTCTCGCTGTGCGCGCGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 537
QY 183 IleAsnIleSerArgAsnArgLysSerThrPheValSerPheThrLysPheGlyAsnLeu 202
```

```

Db 538 ATCTACATTTTCGCGGACCCAGAAAGATAGATGTGCTCCATCTCTGCGACTTTCGGAACGC 597
QY 203 IleAsnProAspLysLeuLeuAspLeuGluLysSerValAspIlePheAlaSerGlyIle 222
Db 598 AATGTGCCAGATCTGTTG-----CTCCAGAGAGTGTATGATCTATATGTGTGCGACT 651
QY 223 SerPheCysGlyProGluTTPAsnPhelGlnAlaGluPheThrAsnAlaAlaSerThrAsn 242
Db 652 GGGTTTGTGTCCTCTCTGCGGATCACTGCTGTTGTTACTGAGGGGCAAGCAAAATAGAT 711
QY 243 ---SerAsnLeuLeuLeuSerTyrGluGluMetLeuGluLysProValGluAsnVal 261
Db 712 CTAGGCCGAGTCTCTTCTTGAATATGAGGAGGTCTTCGAGACCCAGTGAACACCGTT 771
QY 262 LysLysLeuAlaGluPheMetGlyCysGlyPheThrAspAspGluGluLysGlnGlyIle 281
Db 772 AGAGAGCTCGCAACAATTTGTCGGCGCGCATCTCTGACACAGAGGAGAGAGCTGGCAAT 831
QY 282 ValAspGluIleValLysLeuCysSerPheAsnLeuLysAsnGlnValAsnLys 301
Db 832 GTTGGGAGATCGTCAAGCTATGAGCTAGAGAGTCTAGAGGACGACAAAGCCACAAA 891
QY 302 AsnGly---SerSerTyrAsnSerLysIleAspAsnLysHisPheArgLysGlyGlu 320
Db 892 GAAGGCATACAAAGTGTGTACATCAAGTCTCTGCGACGACTCGTATTTTAGGAAGGGGTG 951
QY 321 ValArgAspTrpAlaAsnTyrLeuThrSerGluMetIleLysLysLeuGluThr 338
Db 952 GAAGGGGACTGAGGAAACCATGATCTCTAAGATGGGCGGACATCTGATTCG 1005

RESULT 13
US-10-259-165-443
; Sequence 443, Application US/10259165
; Publication No. US20030135888A1
; GENERAL INFORMATION:
; APPLICANT: Zhu, Tong
; APPLICANT: Wang, Xun
; APPLICANT: Chang, Hur-song
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Moughamer, Todd
; APPLICANT: Provart, Nicholas
; APPLICANT: Rickes, Darrell
; TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE SILENCING
; FILE REFERENCE: 70030-NP
; CURRENT APPLICATION NUMBER: US/10/259,165
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/368,327
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; NUMBER OF SEQ ID NOS: 782
; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta
; SEQ ID NO 443
; LENGTH: 984
; TYPE: DNA
; ORGANISM: Oryza sativa
US-10-259-165-443

Alignment Scores:
Pred. No.: 4.39e-50 Length: 984
Score: 501.50 Matches: 119
Percent Similarity: 50.80% Conservative: 39
Best Local Similarity: 38.26% Mismatches: 106
Query Match: 24.36% Indels: 47
DB: 14 Gaps: 8
```



Db 541 GGCACTCAAGATGGACCTACCAACGAGGCAACAAGCCCAATGCTTTCCAC----- 591  
Qy 208 LeuLeuAspLeuGluLysSerValAspIlePheAlaSerClyIleSerPheCysGlyPro 227  
Db 592 -----GGACCA 597  
Qy 228 GluTrpAsnPheGlnAlaGluPhe---ThrAsnAlaAlaSerThrAsnSerAsnLeuLeu 246  
Db 598 TACTGGGCGCCAGCTGCTGGAGTATTGGCGAGAGCAAGAGGAGGCCACAGAGAGGTCTC 657  
Qy 247 LeuLeuSerTyrGluGluMetLeuGluLysProValGluAsnValLysLysLeuAlaGlu 266  
Db 658 TTCTCTCGCTACGAGAGAGTACACGAGAGACAGAGCGACGCTGAGGAGCTGGCCGAG 717  
Qy 267 PheMetGlyCysGlyPheThrAspAspGluGluGlnGlyIleValAspGluIleVal 286  
Db 718 TTCTATGGCTGCTCCGCTTCTCCGGGAGGAGGAGGAGCGCGCTGCGGAGCCCATCGTC 777  
Qy 287 LysLeuCysSerPheAspAsnLeuLysAsnGlnValAsnLysAsnGlySerTyr 306  
Db 778 GGCTCTGACCTTCGACCTCGACCTTCAGGAGCTTGGAGGTAAACAGGAATGTCCTAAT--- 834  
Qy 307 AsnSerLysIleAspAsnLysHisPhePheArgLysGlyGluValArgAspTrpAlaAsn 326  
Db 835 GATTTCACATCAAAAACGACTGCTTCTACCGAAAGGGGGTGGCTGGAGACTGGGCGCAAT 894  
Qy 327 TyrLeuThrSerGluMetIleLysLysLeuGlu 337  
Db 895 TATTGTGACCTGAGATGGCGGCACACCTGGAT 927

## RESULT 15

US-10-195-144-80  
; Sequence 80, Application US/10195144  
; Publication No. US20030126646A1  
; GENERAL INFORMATION:  
; APPLICANT: BROWN, GREGORY G.  
; APPLICANT: FORMANOVA, NATASA  
; APPLICANT: DENDY, CHARLES  
; APPLICANT: LANDRI, BENOIT S.  
; APPLICANT: CHEUNG, WING  
; APPLICANT: JIN, HUA  
; TITLE OF INVENTION: NUCLEAR FERTILITY RESTORER GENES AND METHODS OF USE IN  
; TITLE OF INVENTION: PLANTS  
; FILE REFERENCE: 16313-0136  
; CURRENT APPLICATION NUMBER: US/10/195,144  
; CURRENT FILING DATE: 2002-10-01  
; PRIOR APPLICATION NUMBER: 60/305,026  
; PRIOR FILING DATE: 2001-07-12  
; PRIOR APPLICATION NUMBER: 60/305,363  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: 60/308,736  
; PRIOR FILING DATE: 2001-07-30  
; NUMBER OF SEQ ID NOS: 128  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 80  
; LENGTH: 2067  
; TYPE: DNA  
; ORGANISM: *Raphanus sativum*  
US-10-195-144-80

## Alignment Scores:

Pred. No.:	2,67e-47	Length:	2067
Score:	483.00	Matches:	107
Percent Similarity:	57.24%	Conservative:	55
Best Local Similarity:	37.81%	Mismatches:	107
Query Match:	23.46%	Indels:	14
DB:	14	Gaps:	7

US-09-854-122-16 (1-391) x US-10-195-144-80 (1-2067)

Qy 56 TyrTrpGlyAspThrMetArgLysGlyPheTrpGlnMetClyTyrLeuValPro 75  
Db 232 TTTTTCGGTTTGGTGAAAATTTGATTTTTCGATCTTGGCAGGAATTTACATTTTCCG 291

Qy 76 GlyIleValAlaPheGluAspAsnPheLysAlaArgGluThrAspIleIleLeuThr 95  
Db 292 GTTTTGGAGGTTTTCNA-----CCGCAAGACACGTGATGATCATCTCTCG 339  
Qy 96 LeuProLysAlaGlyThrTrpThrLysAlaLeuThrPheAlaIleLeuThrArgAsp 115  
Db 340 TACCCCAAAATCAGGCACTACTTGGTCAAGGCCCTCAGAGTCGCTCTGCTTGAGAGATCA 399  
Qy 116 ValAsnHisProSerSerProThrHisProLeuLeuPhePheAsnProHisSerCysVal 135  
Db 400 AAGAACCATCTTCTCAT-----CATCCTCTCTTATATCATATCATATCCCTCATGGCATTA 453  
Qy 136 GlnAsnLeuGluTyrLeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeu 155  
Db 454 CCATTCTTGGAG---ATCGATGTGTACCACGAAAGCTCAAGTCTCTTAACCTAGCCAAAGTTC 510  
Qy 156 AsnGluSerProArgLeuPheAlaGlyHisIleProTyrSerLeuLeuProAlaSerVal 175  
Db 511 TCAGCACTCCGAGGCTGTTCGACTCATATGCTCCACTGCACACGATCCACGAAGCATC 570  
Qy 176 LeuLysSerGlyThrLysIleIleAsnIleSerArgAsnArgLysSerThrPheValSer 195  
Db 571 AAGCACTCTCTTGCAGATTTGTACGTGTGCAGGAACTGTCAGGACACGTTGATCTCG 630  
Qy 196 PheTrpLysPheGly-----AsnLeuIleAsnProAspLysLeuLeuAspLeuGlu 212  
Db 631 TGTGGTGTTCACAGCTGTGCTATATATAAATCGAACCAACAGAGAGATT---CTCGAG 687  
Qy 213 LysSerValAspIlePheAlaSerGlyIleSerPheCysGlyProGluTrpAsnPheGln 232  
Db 688 TCTATGTTTAAACGAGTCTCGCATGGAACCACTATTTTGGACCTTTTGGGATCATCTC 747  
Qy 233 AlaGluPheThrAsnAlaAlaSerThrAsn---SerAsnLeuLeuLeuSerTyrGlu 251  
Db 748 TTGAGTTACTGGAGAGAGCTTGGAGACCCAAAGCATGCTCTTTCATGAGGTATGAG 807  
Qy 252 GluMetLeuGluLysProValGluAsnValLysLysLeuAlaGluPheMetClyCysGly 271  
Db 808 GAGATGAAAGCCGAGCGCTCGTGTGATCAGATCAAGAGACTTGGGAGCTTCTTGGGATGCT 867  
Qy 272 PheThrAspAspGluGluLysGlnGlyIleValAspGluIleValLysLeuCysSerPhe 291  
Db 868 TTTTACTAAGCAAGAGAGATAGTGAATCTGTGGACGGGATCTTGGACCTCTGCTCTCTG 927  
Qy 292 AspAsnLeuLysAsnGlnValAsnLysAsnGlySerSerTyrAsnSerLysIleAsp 311  
Db 928 CGTAATCTGAGCAGTTTGGAGGCTAAACAAACAGGCAATAAACAAT-----GTGGAG 981  
Qy 312 AsnLysHisPhePheArgLysGlyGluValArgAspTrpAlaAsnTyrLeuThrSerGlu 331  
Db 982 CACAAGTTTTTTTCCGTTAAAGAGAGTGGTGACTCGAAAAATTTATCTTACGCTCTGAA 1041  
Qy 332 MetIleLys 334  
Db 1042 ATGGAGAAA 1050

Search completed: March 20, 2004, 06:32:40  
Job time : 540 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2004, 01:36:53 ; Search time 17 Seconds  
(without alignments)  
2212.405 Million cell updates/sec

Title: US-09-854-122-16  
Perfect score: 2059  
Sequence: 1 RFGTGIACVACPLIMAGIL.....CLLSYFVNNFILLKKKK 391

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	616	29.9	329	2 T47448	sulfotransferase-1
2	602	29.2	323	2 T47447	sulfotransferase-1
3	567	27.5	346	2 E86319	probable flavonol
4	562.5	27.3	338	2 A96769	protein flavonol s
5	558.5	27.1	350	2 H96768	protein flavonol s
6	546	26.5	325	2 T07833	probable steroid s
7	545	26.5	331	2 E84451	probable steroid s
8	544.5	26.4	326	2 F86407	probable sulfotran
9	537.5	26.1	326	2 A84452	probable steroid s
10	529.5	25.7	323	2 T07831	probable steroid s
11	527.5	25.6	302	2 S69188	probable flavonol
12	527	25.6	324	2 B84452	probable steroid s
13	525.5	25.5	333	2 A84523	probable steroid s
14	520.5	24.8	320	2 A40316	flavonol 4'-sulfot
15	500.5	24.3	324	2 T07832	probable steroid s
16	486.5	23.6	314	2 T06012	hypothetical prote
17	455	22.1	311	2 B40216	flavonol 3'-sulfot
18	309	15.0	295	2 S29045	estrone sulfotransf
19	278.5	13.5	301	2 JW0078	amine sulfotransf
20	258	12.5	299	2 JE0196	sulfotransferase (
21	255.5	12.4	304	2 A49098	N-hydroxyarylamine
22	255	12.4	294	2 JC2229	estrogen sulfotran
23	253.5	12.3	298	2 S28183	aryl sulfotransfer
24	253	12.3	295	2 A41930	estrone sulfotrans
25	252	12.2	296	2 JC7921	cytosolic sulfotra
26	250	12.1	295	2 I56606	estrogen sulfotran
27	250	12.1	295	2 I73679	estrogen sulfotran
28	245.5	11.9	295	2 I53296	testis-specific es
29	245	11.9	291	2 S10329	aryl sulfotransfer

30	243.5	11.8	285	1 I38548	alcohol sulfotrans
31	243.5	11.8	296	2 JC5885	thyroid hormone su
32	239.5	11.6	286	2 A44011	adrenocortical est
33	234.5	11.4	286	2 JE0152	alcohol sulfotrans
34	233	11.3	296	2 JC7282	hydroxyarylamine s
35	232	11.3	295	2 JE0197	phenol sulfotransf
36	232	11.3	299	2 JC5884	thyroid hormone su
37	229.5	11.1	295	2 JC2523	aryl sulfotransfer
38	229	11.1	296	2 JC7283	hydroxyarylamine s
39	228.5	11.1	284	2 JC5000	aryl sulfotransfer
40	226.5	11.0	295	2 G01843	aryl sulfotransfer
41	225.5	11.0	295	2 JC5249	aryl sulfotransfer
42	225	10.9	295	2 S52791	aryl sulfotransfer
43	223.5	10.9	295	2 S52399	aryl sulfotransfer
44	223.5	10.9	295	2 JC5248	aryl sulfotransfer
45	223.5	10.9	295	2 S52794	aryl sulfotransfer

ALIGNMENTS

RESULT 1

T47448  
sulfotransferase-like protein - Arabidopsis thaliana  
N:Alternate names: Protein T14D3.20  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 02-Sep-2000  
C:Accession: T47448  
R:Jordan, N.; Bangert, S.; Wiedelmann, R.; Voss, H.; Unseld, M.; Mewes, H.W.; Lemcke, K.  
Submitted to the Protein Sequence Database, February 2000  
A:Reference number: 224467  
A:Accession: T47448  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-329 <JOK>  
A:Cross-references: EMBL:AL138649  
A:Experimental source: cultivar Columbia; BAC clone T14D3  
C:Genetics:  
A:Map position: 3  
A:Note: T14D3.20  
C:Superfamily: alcohol sulfotransferase

Query Match	29.9%	Score	616	DB	2	Length	329
Best Local Similarity	41.2%	Pred. No.	1.3e-36				
Matches	132	Conservative	62	Mismatches	100	Indels	26
Gaps	9						
QY	29	KNEQEKEDSKYKRYREIVSSLPN-DYWGDTWRLYKGFQWQGYLYPGIMAFEDNFKAR	87				
Db	11	RNDELSEES-----KTLISLSPSDKXNSTGVNVCKYGGCWYTPPIQGVLFNFQNFKPQ	63				
QY	88	ETDIITLTPKAGTTWTKALTFAITRDVNHPS-SPTHPLLFNPHSCVQNLK-ELYMGR	145				
Db	64	DTDIIVASFFKCGITWTKALTFALVRRS-KHPSHDDHFLPSDNPVLSPLSLEMYLYLCS	122				
QY	146	ENTMPDLMLESPLPAGHTPYSLLPASVLKSGTKIINISRNKSTFVSFWKF-----	199				
Db	123	EN--PDLTKFSSSSRLFTSHPSHTLOEGLKGSTCKIVMSRNVDKTLVSVWHFCKQT	180				
QY	200	GNLNPDKLLDLEKSVDFISAGISFCGPEWFOAEFTNAASTNSN-ILLLSYSBEMLEKP	257				
Db	181	DNIIIS-----SVETTFEMFCRGVNFYFPFDHVLVSRGSLSDPNHVLPMKFEEMKEEP	235				
QY	258	VENVKLAEPFGCGTDDERKQGVDEIVKLCSPDNLNKQOVKNNGSSYNKIDNKHFR	317				
Db	236	REQIKRLAEFLGCLFTKEEESGLVDEIIDLCRLNLSLEINKTKLHSTGRENKTFPR	295				
QY	318	KGEVDMANLYLTSEMIXKLE	337				
Db	296	KGEVDMANLYLTPEMENKID	315				

RESULT 2  
T47447

sulfotransferase-like protein - Arabidopsis thaliana

N/Alternate names: protein T14D3.10  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 20-Apr-2000 #sequence\_revision 02-Mar-2000 #text\_change 02-Sep-2000  
C/Accession: T47447  
R/Jordan, N.; Bangert, S.; Wiedemann, R.; Voss, H.; Unsel, M.; Mewes, H.W.; Lemcke, K.  
submitted to the Protein Sequence Database, February 2000  
A/Reference number: Z24467  
A/Accession: T47447  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-323 <JOB>  
A/Cross-references: EMBL:AL138649  
A/Experimental source: cultivar Columbia; BAC clone T14D3  
C/Genetics:  
A/Map position: 3  
A/Note: T14D3.10  
C/Superfamily: alcohol sulfotransferase

Query Match 29.2%; Score 602; DB 2; Length 323;  
Best Local Similarity 42.6%; Pred. No. 1.3e-35;  
Matches 127; Conservative 51; Mismatches 112; Indels 8; Gaps 6;  
QY 45 REIVSLPSN-DYWGDTVRLYKGFQMGVLPVGINAFEDNFKARETDIIITLTPKAGTTW 103  
DB 15 KTLISLPSDKDFTGKTICKYQGCWYTHNVLQAVLNFKSPKPDTDIIIVASFPKCGTTW 74  
QY 104 KTAITFAITLDRVNHPSPTPLLPNHSVCQNLLEY-LYMGRENTMPDLDMLNESPLRF 162  
DB 75 LKALTFALLHRSKQSHDDHLLNPNHVLVPEIDLYRSEN--PDLTKFSSPLRF 132  
QY 163 AGHIPVSLPASVLKSGTKIINISNRKSTFTVSWFKGNLINPDK--LLDLEKSVDPFAS 220  
DB 133 STHVSHTLQEGLGKSTCKIVISNRKDTLVSYWHFTKQTDKLISSPEDTFENFCR 192  
QY 221 GSFGCPENFQAEFTNAASTNSN-LLLSYEEMLEKPVNVKLAEFWCGCTDDDEKQ 279  
DB 193 GVSIFGPFVDHVLVYRSGLEDPNHVLFMKFEEMKAEPRDQIKFAEFLGCGPFTKEEES 252  
QY 280 GVDEIVKLCSPDNKQVKNKNGSYNSKIDNKHFFRKGVRDWNATLTSEMINKLE 337  
DB 253 GSVDEIIDLCSRLNLSLEINKTG-KLNSGRENKMFRRKGEVGDWKNVLTPEMENKID 309

## RESULT 3

E86319  
probable flavonol sulfotransferase [imported] - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C/Accession: E86319  
R/Theologas, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Huizar, L.  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A/Authors: Salzborg, S.B.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A/Reference number: A86141; MUID:21016719; PMID:11130712  
A/Accession: E86319  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-346 <STO>  
A/Cross-references: GB:AE005172; NID:99795597; PIDN:AAF98415.1; GSPDB:GN00141  
C/Genetics:  
A/Map position: 1  
C/Superfamily: alcohol sulfotransferase

Query Match 27.5%; Score 567; DB 2; Length 346;  
Best Local Similarity 39.6%; Pred. No. 4.4e-33;  
Matches 137; Conservative 56; Mismatches 123; Indels 30; Gaps 11;

QY 23 EKCFGSKNEQKEEDSKMYKYREIVSVSLPSNDYK--GDTMRLYKGFQMGVLPVGINAF 80  
DB 17 ELASSPSEFEKQ-----KHYQEIATLPHDGRPKDPFVEYCGHWLQPLEGLLHA 71  
QY 81 ENFKARETDIIITLTPKAGTTWTKALTFAILTR--DVNHPSSTHPLLPNHSVCQVN 137  
DB 72 QKFFKARPNDPFCVSPKGTITLTKALTFANRSKFDVS-----TNPLLRNPHFVPY 126  
QY 138 LEYLVMGRENTMPDLDML-NESPRIFAGHIPVSLPASVLKSGTKIINISNRKSTFTVSV 196  
DB 127 IBDP-----PFPFSDVLKDEGNTLFTSHIYDLPFESVVKSGCKIVIRWDPKDTFVSM 182  
QY 197 WKFGNLINPDK--LLDLEKSVDFASGISFCGPEWNFQAEFTNAASTN-SNLLLSYSEEM 253  
DB 183 WFAFKERSQCGPVVSEBAPDKYCOGLSAYGYLDHVLGYWKAYQANPDQILFKYETM 242  
QY 254 LEKPVNVKLAEFWCGCTDDDEKQIVDEIVKLCSPDNKQVKNKNGSYNSK--I 310  
DB 243 RADPLPYVVRLAEFMGYGTKEEBEGNVVVKVLCSPFETLKNLEANKGEKDRDPAYV 302  
QY 311 DNKHFFRKGVRDWNATLTSEMINKLETAGKINESEK--HLLSNKN 354  
DB 303 ANSAYFRKCKVGDWQNVLTPEVARID--GLMEKFKGTGFLSSKS 346  
RESULT 4  
A96769  
protein flavonol sulfotransferase F2P9.3 [imported] - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001  
C/Accession: A96769  
R/Theologas, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Huizar, L.  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A/Authors: Salzborg, S.B.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A/Reference number: A86141; MUID:21016719; PMID:11130712  
A/Accession: A96769  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-338 <STO>  
A/Cross-references: GB:AE005173; NID:97109462; PIDN:AAF36726.1; GSPDB:GN00141  
C/Genetics:  
A/Map position: 1  
C/Superfamily: alcohol sulfotransferase

Query Match 27.3%; Score 562.5; DB 2; Length 338;  
Best Local Similarity 38.4%; Pred. No. 8.9e-33;  
Matches 129; Conservative 59; Mismatches 123; Indels 25; Gaps 10;  
QY 33 EKEDSKMYKYREIVSVSLPSNDYK--GDTMRLYKGFQMGVLPVGINAFEDNFKARETD 90  
DB 14 ELTBEFTQKXQDFIATLPKSGWRPDEILTQYGGHWQECLEGLFHAHDFEARPTD 73  
QY 91 IILTLPKAGTTWTKALTFAILTRDVNHPSPTPLLPNHSVCQNLLEYVMGRENTMP 150  
DB 74 FLVCSYPTGTITLTKALTYAVNR--SRDYDAANPLLRNPHFVPVVEIDFA----FYP 127  
QY 151 DLDMLNESPR-LPAGHIPVSLPASVLKSGTKIINISNRKSTFTVSWFKGNLINPDK-- 207  
DB 128 TVDVLQDRKPLFTSHLPNGLLPDSIVNSCKVYIWRDPKDTFISMTWTF--LHKEKSQ 184  
QY 208 ---LLDLEKSVDFISGISFCGPEWNFQAEFTNAASTN-SNLLLSYSEEMLEKPVENVK 263  
DB 185 EGQLASLESDSPMFCGLSVGYLDHVLGYWKAYQENPDRIILFLRYETVRANPLPFVKR 244







11 REEBEKPSEEFKILISSIPWIDVLGNKLFNYEGWYSEDLQSPINHTGFOQETDII 70  
93 LTTLPKAGTTWKALTFAILLTRDVNHPSPHPLFFNPHSCVONLEY-LYMGRENTMPD 151  
71 LASFYKSGTTWKALTFAIVORSKHSLEDHQHLLHNPHIEIVPNLELDLYL--KSKRPD 128  
152 LDML---NESPRIFAGHIPYSLLPASVLKSGTKIINISNRKSTFVSFWKF--GNLNP 205  
129 LTKFLSSSSSPRIFSTHMSLDLPLOVPLKNCILVYVCRNVKDVWVSVYFRQSKITR 188  
206 DKLLDLEKSVDFIFASGISFCGPEWFOAEFTNAASTN--SNLLLSYEEMLEKPVNKKL 264  
189 AEDYSLEAIFESFCNGVTLHGPFWDHALSYWRGSLDPKHFVRYEDLKAEPPTQVKRL 248  
265 AEFMCGGTTDDEKOGIVDEIVKLCSPDNLNQVKNKNGSSYKINDKNHFFRKGEVRDW 324  
249 AEFLOCPPTKBEEDSGVDKILELCSLNSLRSVEINKTRTS--SRVDFKSYFRKGQVGDW 306  
325 ANYLTSEMIMKLE 337  
307 KSYMTPENVDKID 319

## RESULT 14

A40216  
flavonol 4'-sulfotransferase - Flaveria chloraefolia  
C:Species: Flaveria chloraefolia  
C:Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 20-Sep-1999  
C:Accession: A40216  
R:Varin, L.; Deluca, V.; Ibrahim, R.K.; Brisson, N.  
Proc. Natl. Acad. Sci. U.S.A. 89, 1286-1290, 1992  
A:Title: Molecular characterization of two plant flavonol sulfotransferases.  
A:Reference number: A40216; PMID:92159034; PMID:1741382  
A:Accession: A40216  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-320 <VAR>  
A:Experimental source: terminal bud  
A>Note: sequence inconsistent with the nucleotide translation  
A>Note: sequence extracted from NCBI backbone (NCBIN:82216, NCBIIP:82217)  
C:Superfamily: alcohol sulfotransferase

Query Match 24.8%; Score 510.5; DB 2; Length 320;  
Best Local Similarity 37.6%; Pred. No. 4.3e-29;  
Matches 120; Conservative 58; Mismatches 110; Indels 31; Gaps 9;  
QY 35 EDSRMVRYREIVSSLP----SNDYWGDTMLYKGFQWQGYLVPGIMAFEDNFKARETD 90  
DB 2 ETTKQFESMAEMIKLPQHTCSSLGRITLYKYQDFWGLQNNIEGAILAQOQSFARPD 61  
QY 91 IILTLPKAGTTWKALTFAILLTRDVNHPSPHPLFFNPHSCVONLEYLYMGRENTMP 150  
DB 62 VFLCSVPKSGTTWLKALAYAVTR--KDEFTSPLTNIHNCIPIYI-----K 109  
QY 151 LDML--NESPRIF--AGHIPYSLLPASVLKSGTKIINISNRKSTFVSFWKFGNLLNP 205  
DB 110 DLAKIVDNQNSCFTPMATHPVLPKSLALNCKVYIYRNKIDVIVSYFHGREITK 169  
QY 206 DKLLD--LEKSVDFIFASGISFCGPEWFOAEFTNAASTNLL--LSYEEMLEKPVNKK 262  
DB 170 LPLEADPFEEAFDEFYHGISQFGPYWDHLLGYKWSLERPEVILFLKYEDVKDPTSNVY 229  
QY 263 KLAFFMCGGTTDDEKOGIVDEIVKLCSPDNLNQVKNKNGSSYKINDKNHFFRK 318  
DB 230 RLAEFIGYPTFEEKEGVIESIKLCSFENLNLNLEVNKSG---NSKGFLEPNRLYFRK 286  
QY 319 GEVRDWNALYLTSEMIMKLE 337  
DB 287 AKOGDWKNYFTDENTEKID 305

## RESULT 15

T07832  
probable steroid sulfotransferase (EC 2.8.2.15) 2 - rape  
C:Species: Brassica napus (rape)  
C:Date: 14-May-1999 #sequence\_revision 14-May-1999 #text\_change 08-Oct-1999  
C:Accession: T07832  
R:Richard, M.; Nicolle, L.; Varin, L.  
submitted to the EMBL Data Library, April 1997  
A:Reference number: Z16161  
A:Accession: T07832  
A>Status: translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-324 <RIC>  
A:Cross-references: EMBL:AF000306; NID:g3420005; PIDN:AAC63112.1; PID:g3420006  
C:Superfamily: alcohol sulfotransferase  
C:Keywords: sulfotransferase

Query Match 24.3%; Score 500.5; DB 2; Length 324;  
Best Local Similarity 38.0%; Pred. No. 2.3e-28;  
Matches 117; Conservative 58; Mismatches 114; Indels 19; Gaps 10;  
QY 37 DSRMRYREIVSSLPNDYWGDTMLYKGFQWQGYLVPGIMAFEDNFKARETDIILT 95  
DB 13 DENLTOKTKDLISLSPSEKGLVQCMYQQRWHTQALLOGLTCQKHFEAKDSIILVT 72  
QY 96 LPKAGTTWKALTFAILLTRDVNHPSPHPLFFNPHSCVON--LEYLYMGRE--NTMPDLD 153  
DB 73 NPKSGTITWLKALVFAI----INRHKFPVYSVAIL---SCYQSALLVPFLGRSILRSDFD 125  
QY 154 MLN--ESPRIFAGHIPYSLLPASVLKSGTKIINISNRKSTFVSFWKFGNLLNPDKLLD-- 210  
DB 126 FSQSSPRLMNTHTSHLSLPESVYKSSCKIVYCCRNPKDMFVSLMHFGKGLAPEETADYP 185  
QY 211 LEKSVDFIFASGISFCGPEWFOAEFTNAASTNSN--LILLSYEEMLEKPVNKKLAFFMG 269  
DB 186 IEKAEAFCCGKTIQGGFWHDVLEWYASLENFKNVLFVSYBEPKXKTKGTETIKRIAEFLG 245  
QY 270 CGFTDDEKOGIVDEIVKLCSPDNLNQVKNKNGSSYKINDKNHFFRKGEVRDWNALYLT 329  
DB 246 CGLVGBEE---VRAIVKLCSPESLSLEVNREG-KLPSCMETRAFFRKGEVGGWRDILT 300  
QY 330 SEMIKKLE 337  
DB 301 ESLEAVID 308

Search completed: March 20, 2004, 01:39:46  
Job time : 17 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2004, 01:35:33 ; Search time 11 Seconds  
(without alignments)  
1850.857 Million cell updates/sec

Title: US-09-854-122-16

Perfect score: 2059

Sequence: 1 TRGITGIACVPLIMAGIL.....CILLSYFVINPILKKKKK 391

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	537.5	26.1	326	1	FSTL_ARATH
2	510.5	24.8	320	1	P4STL_FLACH
3	491.5	23.9	312	1	P3ST_FLABI
4	473	23.0	309	1	FSTL_FLABI
5	455	22.1	311	1	F3ST_FLACH
6	309	15.0	295	1	SUOE_BOVIN
7	304	14.8	302	1	S1C2_HUMAN
8	255.5	12.4	304	1	SUAC_RAT
9	255	12.3	294	1	STOE_HUMAN
10	253	12.3	295	1	SUO3_RAT
11	252	12.2	295	1	SUO1_RAT
12	252	12.2	295	1	SUO2_RAT
13	250	12.1	295	1	SUO6_RAT
14	248.5	12.1	291	1	SUAR_MOUSE
15	248.5	12.1	295	1	SUOT_MOUSE
16	245	11.9	291	1	SUAR_RAT
17	243.5	11.8	284	1	SUHA_HUMAN
18	239.5	11.6	296	1	SUOE_CAVPO
19	238	11.6	296	1	STK2_RAT
20	233	11.3	296	1	STK1_RAT
21	232	11.3	299	1	SUDY_RAT
22	229	11.1	284	1	SUHA_MACFA
23	228.5	11.1	294	1	SUPP_BOVIN
24	228	11.1	296	1	S1C1_RABIT
25	225.5	11.0	295	1	SUF2_HUMAN
26	224	10.9	296	1	S1C1_HUMAN
27	223.5	10.9	295	1	SUF1_HUMAN
28	222.5	10.8	295	1	SUPM_HUMAN
29	221	10.7	295	1	SUPP_MACFA
30	213	10.3	285	1	SUH2_MOUSE
31	212	10.3	285	1	SUH2_RAT
32	211.5	10.3	283	1	SUHS_RAT
33	211.5	10.3	286	1	SUHB_CAVPO

34	208.5	10.1	286	1	SUHA_CAVPO
35	205.5	10.0	283	1	SUHA_RAT
36	205.5	10.0	285	1	SUH3_RAT
37	201	9.8	285	1	SUH1_MOUSE
38	187	9.1	284	1	S4A1_HUMAN
39	182	8.8	284	1	S4A1_MOUSE
40	119	5.8	1628	1	YATE_SCHPO
41	119	5.8	1726	1	MSPI_PLAFC
42	119	5.8	1726	1	MSPI_PLAPP
43	118.5	5.8	944	1	UVRA_MYCFU
44	115.5	5.6	1701	1	MSPI_PLAPP
45	115.5	5.6	1701	1	MSPI_PLAFM

#### ALIGNMENTS

#### RESULT 1

FSTL_ARATH	STANDARD;	PRT;	326 AA.
AC	P52839; O9SJW2;		
DT	01-OCT-1996 (Rel. 34, Created)		
DT	28-FEB-2003 (Rel. 41, Last sequence update)		
DT	15-MAR-2004 (Rel. 43, Last annotation update)		
DE	Flavonol sulfoltransferase-like (EC 2.8.2.-) (RaRO47).		
GN	AT2G3760 OR F19B11.21.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;		
OC	euroside II; Brassicales; Brassicaceae; Arabidopsis.		
OX	NCBI_TaxID=3702;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CV. Columbia;		
RX	MEDLINE=96270377; PubMed=8639757;		
RA	Lacomme C., Roby D.;		
RT	"Molecular cloning of a sulfoltransferase in Arabidopsis thaliana and regulation during development and in response to infection with pathogenic bacteria.";		
RT	Plant Mol. Biol. 30:995-1008(1996).		
RL	[2]		
RN	SEQUENCE FROM N.A.		
RP	STRAIN=CV. Columbia;		
RC	MEDLINE=20083487; PubMed=10617197;		
RA	Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D., Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V., Buehl C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L., Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L., Tallon L.J., Gill J.B., Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Sonerville C.R., Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M., Venter J.C.;		
RT	"Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.";		
RT	Nature 402:761-768(1999).		
RL	[3]		
RN	SEQUENCE FROM N.A.		
RP	STRAIN=CV. Columbia;		
RC	MEDLINE=22954850; PubMed=14593172;		
RX	Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J., Kuroki M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F., Southwick A.M., Wu H.C., Kim C.J., Lam B., Sakano H., Wu T., Yu G., Karlin-Newmann G., Liu S.X., Lam B., Chang C.H., Lee J.M., Toriumi M.J., Miranda M., Quach H.L., Tripp M., Chang C.H., Akiyama K., Ansari Y., Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Brooks S.Y., Carninci P., Arakawa T., Ban H., Banno F., Bowser L., Gurjal M., Hansen N.F., Chao Q., Choy N., Enju A., Goldsmith A.D., Guan V.W., Iida K., Karnes M., Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Jones T., Kawai J., Khan S., Koesema B., Ishida J., Jiang P.X., Jones T., Sakurai T., Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T., Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y., Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;		
RT	"Empirical analysis of transcriptional activity in the Arabidopsis genome.";		

```

Science 302:842-846(2003).
[4]
RN SEQUENCE OF 1-93 FROM N.A.
RP STRAIN:cv. Columbia;
RC Lacomme C., Roby D.;
RL Submitted (Apr-1993) to the EMBL/GenBank/DBJ databases.
RR Submitted (Apr-1993) to the plant sulfotransferase family.
CC -!- SIMILARITY: Belongs to the plant sulfotransferase family.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to frameshifts
CC in positions 300, 313, and 317.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; 246823; CA86850.1; ALT_FRAME.
DR EMBL; AC008636; AAD20078.1; -.
DR EMBL; AF375458; AAK53042.1; -.
DR EMBL; AY113050; AAM47358.1; -.
DR EMBL; Z23001; CA880546.1; ALT_INIT.
DR PIR; A84452; A84452.
DR HSP; P50224; ICGM.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase; 1.
DR ProDom; PD001218; Sulfotransferase; 1.
DR Transferrase.
DR KW CONFLICT 85 93 VPALLNRHK -> ALPSTVPV (IN REF. 4).
DR SQ SEQUENCE 326 AA; 37139 MW; 085069A2EBD4E10A CRC64;
-----
Query Match 26.1%; Score 537.5; DB 1; Length 326;
Best Local Similarity 39.7%; Pred. No. 4.2e-31;
Matches 123; Conservative 54; Mismatches 112; Indels 21; Gaps 9;
-----
QY 37 DSKMYKRYREIVSSLPNDYWGDMRLYKGMQWGLVPGIMAFEDNFKARETDIILT 95
DB 13 DEDLTQETRALISLPKRGWLVSEIYFQGLWHTQAILQILICQKFEAKDSIILVT 72
QY 96 LPAKAGTWTWKALTFTALTDVNH---PSSPHLLPFNPHSCVQNLVLYMGRENTWPD 151
DB 73 NPKSGTTLKALVFALLNR---HKFPVSSGNHPLLVNTHLLVFLFGLVY---ESPD 125
QY 152 LDMNLN-ESPRLPFAGHPYSLLPASVLKSGTKIINISRNKSTFVSFKFGLNLPDKLLD 210
DB 126 FDFSSLPRLNTHSHLSLPESVKSCKIVYCCRNPKDMFVSLMHFGKLADEETAD 185
QY 211 --LEKSVDFIFASGIFSGCPFWNFOAFTNAASTNSN-LILLLSYEMLKPKVENYKLAEF 267
DB 186 YPIEKAVEAFCEGFTGGFPWPHILEYVYASRENPNKVLFTVYBELKXQTEVEMKRIAEF 245
QY 268 MCGGFTDDBEKQIGVDEIVKLCSDNLKQVKNQNGSSYNKIDNKHFFRKGEVRDWANY 327
DB 246 LECGFIEEE---VREIVKLCFSFSLNLEVNKESGLPNG-IETKTFPRKGEIGGWRDT 300
QY 328 LTSEMICKLE 337
DB 301 LSESLAEED 310
-----
RESULT 2
FAST_FLACH STANDARD; PRT; 320 AA.
AC P52837;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Flavonol 4'-sulfotransferase (EC 2.8.2.-) (F4-ST).
OS Flaveria chloraefolia.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC campanulids; Asterales; Asteraceae; Asteroidae; Tageteae; Flaveria.
-----
NCBI_TaxID=4228;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=92159034; PubMed=1741382;
RA Varin L., Deluca V., Ibrahim R.K., Brisson N.;
RT "Molecular characterization of two plant flavonol sulfotransferases.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:1286-1290(1992).
RN [2]
RP PAPS-BINDING SITE.
RX MEDLINE=95279378; PubMed=7759495;
RA Varin L., Marsolais F., Brisson N.;
RT "Chimeric flavonol sulfotransferases define a domain responsible for
RT substrate and position specificities.";
RL J. Biol. Chem. 270:12498-12502(1995).
CC -!- FUNCTION: Transfers sulfate group into flavonol position 4'.
CC May play a role in auxin transport.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- TISSUE SPECIFICITY: Highest in shoot tips and lowest in mature
CC leaves and roots (By similarity).
CC -!- SIMILARITY: Belongs to the plant sulfotransferase family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M84136; AAA33343.1; -.
DR HSP; P50224; ICGM.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase; 1.
DR ProDom; PD001218; Sulfotransferase; 1.
DR Transferrase.
DR KW BINDING 148 163 PAPS-BINDING SITE.
DR SQ SEQUENCE 320 AA; 37255 MW; 6B263659F6CCBC0 CRC64;
-----
Query Match 24.8%; Score 510.5; DB 1; Length 320;
Best Local Similarity 37.6%; Pred. No. 3.4e-29;
Matches 120; Conservative 58; Mismatches 110; Indels 31; Gaps 9;
-----
QY 35 EESKMYKRYREIVSSLP---SNDYWGDMRLYKGMQWGLVPGIMAFEDNFKARETD 90
DB 2 ETTKTOFESMAEIKKLPQHTCSSLGRITLYKIQDFWGLQNNIEGAILAQSFARPPD 61
QY 91 IILTTPKAGTWTWKALTFTALTDVNHPSSTPHLLPFNPHSCVQNLVLYMGRENTWPD 150
DB 62 VFLCSYKSGTTLKALAVAVTRE--KFDEFTSPLLTINPHNCIPYIE-----K 109
QY 151 LDMLN--NESPRLP---AGHIPYSLLPASVLKSGTKIINISRNKSTFVSFKFGLNLP 205
DB 110 DLKXIVENQNSCFTPMATHPYHVLKPSILALNCQWYIYRNKIDVIVSFHFGEITK 169
QY 206 DKLLD--LEKSVDFIFASGIFSGCPFWNFOAFTNAASTNSNLL-LSYEMLKPKVENYK 262
DB 170 LPLEDAFFEAEDFYHGISQFGYWDHLLGYWKASLERPEVILFKYEDVKDPTSNVK 229
QY 263 KLAFFMCGGFTDDBEKQIGVDEIVKLCSDNLKQVKNQNGSSYNK-----IDNKHFFRK 318
DB 230 RLAEFTGYPTFEEKEGVIESIKLCSFENLSNLEVNKSG---NSKGLPIENRLYFRK 286
QY 319 GEVRDWANYLTSEMICKLE 337
DB 287 AKGDWKNYFTDENTEXID 305
-----
RESULT 3
FAST_FLABI STANDARD; PRT; 312 AA.
AC P52835;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
```





OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
 CC campanulids; Asterales; Asteraceae; Asteroideae; Tageteae; Flaveria.  
 CX NCBI\_TaxID=4228;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92159034; PubMed=1741382;  
 RA Varin L., Deluca V., Ibrahim R.K., Brisson N.;  
 RT "Molecular characterization of two plant flavonol sulfotransferases.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:1286-1290(1992).  
 RN [2]  
 RP SEQUENCE OF 25-38 AND 159-190.  
 RX MEDLINE=92112914; PubMed=1309801;  
 RA Varin L., Ibrahim R.K.;  
 RT "Novel flavonol 3-sulfotransferase. Purification, kinetic properties,  
 RT and partial amino acid sequence.";  
 RL J. Biol. Chem. 267:1858-1863(1992).  
 RN [3]  
 RP MUTAGENESIS OF LYS-59; GLU-101; ARG-276; GLY-281 AND LYS-284.  
 RX MEDLINE=96107199; PubMed=8530475;  
 RA Marsolais F., Varin L.;  
 RT "Identification of amino acid residues critical for catalysis and  
 RT cosubstrate binding in the flavonol 3-sulfotransferase.";  
 RL J. Biol. Chem. 270:30458-30463(1995).  
 RN [4]  
 RP PAPS-BINDING SITE.  
 RX MEDLINE=95279378; PubMed=7759495;  
 RA Varin L., Marsolais F., Brisson N.;  
 RT "Chimeric flavonol sulfotransferases define a domain responsible for  
 RT substrate and position specificities.";  
 RL J. Biol. Chem. 270:12498-12502(1995).  
 CC -1- FUNCTION: CATALYZES THE SULFATE CONJUGATION OF QUERCETIN,  
 CC RHAMNETIN AND ISORHAMNETIN BUT NOT KAEMPFEROL. O-SULFATION OF  
 CC POSITION 3 OF FLAVONOL. MAY PLAY A ROLE IN AUXIN TRANSPORT.  
 CC OPTIMAL PH 6.0 AND 8.5.  
 CC -1- SUBUNIT: Monomer (potential).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- TISSUE SPECIFICITY: Highest in shoot tips and lowest in mature  
 CC leaves and roots (by similarity).  
 CC -1- PM: The N-terminus is blocked.  
 CC -1- SIMILARITY: Belongs to the plant sulfotransferase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; M84135; AAA33342.1; -.  
 DR PIR; B40216; B40216.  
 DR HSP; P50224; 1CJM.  
 DR InterPro; IPR000863; Sulfotransferase.  
 DR Pfam; PF00685; Sulfotransferase; 1.  
 DR ProDom; PD001218; Sulfotransferase; 1.  
 KW Transferase.  
 FT BINDING 137 152 PAPS-BINDING SITE.  
 FT MUTAGEN 59 59 K->A: REDUCTION OF ACTIVITY.  
 FT MUTAGEN 101 101 E->K: NO EFFECT.  
 FT MUTAGEN 276 276 R->K: REDUCTION OF ACTIVITY.  
 FT MUTAGEN 276 276 R->E: REDUCTION OF ACTIVITY.  
 FT MUTAGEN 276 276 R->A: REDUCTION OF ACTIVITY.  
 FT MUTAGEN 281 281 G->A: REDUCTION OF ACTIVITY.  
 FT MUTAGEN 284 284 K->G: REDUCTION OF ACTIVITY.  
 SQ SEQUENCE 311 AA; 36484 MW; F51678902DB7D9E0 CRC64;  
 Query Match 22.1%; Score 455; DB 1; Length 311;  
 Best Local Similarity 37.8%; Pred. No. 2.9e-25;  
 Matches 110; Conservative 43; Mismatches 110; Indels 28; Gaps 8;  
 QY 50 TMLRYKGFQWGVLPVCIAPFENFKARSTDIILTLPRACWTWKATFTAILPRDYNHP.119  
 DB 21 TLKYKDANWHQBLEGRILSEQKFAHPNDVFLASYPKSGTTLWKA-WICILTRKFPDD 79

QY 120 SSTPTLLFFNPHSCVQNLVLYMGRENTMPDLMLAESPR-----LFAGHIPYSLLPAS 174  
 DB 80 S--TSPLLTTPHDCIPLE-----KDLKIQENQNSLYTISTHFKSLPES 127  
 QY 175 VLKSGTKIINISRNKSTFTVFWKFGNLINPKL-----LDLEKSVDFPASISFCGPEWN 230  
 DB 128 ARTSNCKIIVIVRNKMDVIVSYTHF--LRQIVKLSVEEAPPEAFDFCQGISCGPYWE 185  
 QY 231 FQAEFTNRAASTNSNL-LLLSYEEMLEKPVENVYKLABFMGCGFTDDEBKQIGVDEIVKLC 289  
 DB 186 HIKGYWKASLEKPEIFELKYEDMKKDPVPSVKKLADFIHPFTPKSEAGVIEDIVKLC 245  
 QY 290 SFDNLKNQOVNKNQSSYNSK---IDNKHFRKGEVRDWMANYLTSEMKKLE 337  
 DB 246 SPEKLSLEVNKSGMHRPEEAHSIENRLYFRKKGDKGDKWKNYFTDENTQKID 296  
 RESULT 6  
 SUCR BOVIN  
 ID \_SUOE\_BOVIN STANDARD; PRT; 295 AA.  
 AC P19217;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Estrogen sulfotransferase (EC 2.8.2.4) (Sulfotransferase, estrogen-  
 DE preferring) (STIE1).  
 GN STE OR SUL1E1 OR OST.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 OX NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=90147277; PubMed=3271383;  
 RA Nash A.R., Glenn W.K., Moore S.S., Kerr J., Thompson A.R.,  
 RA Thompson E.O.P.;  
 RT "Oestrogen sulfotransferase: molecular cloning and sequencing of cDNA  
 RT for the bovine placental enzyme.";  
 RL Aust. J. Biol. Sci. 41:507-516(1988).  
 RN [2]  
 RP PARTIAL SEQUENCE.  
 RC TISSUE=Placenta;  
 RX MEDLINE=90104087; PubMed=3270501;  
 RA Moore S.S., Thompson E.O.P., Nash A.R.;  
 RT "Oestrogen sulfotransferase: Isolation of a high specific activity  
 RT species from bovine placenta.";  
 RL Aust. J. Biol. Sci. 41:333-341(1988).  
 RN [3]  
 RP SEQUENCE OF 146-160 AND 206-220, AND CHARACTERIZATION.  
 RC TISSUE=Placenta;  
 RX MEDLINE=91152101; PubMed=1900200;  
 RA Adams J.B.;  
 RT "Enzymic synthesis of steroid sulphates. XVII. On the structure of  
 RT bovine estrogen sulphotransferase.";  
 RL Biochim. Biophys. Acta 1076:282-288(1991).  
 CC -1- FUNCTION: May control the level of the estrogen receptor by  
 CC sulfonylating free estradiol.  
 CC -1- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + estrone =  
 CC adenosine 3',5'-bisphosphate + estrone 3-sulfate.  
 CC -1- SUBUNIT: Homodimer.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- INDUCTION: By progesterone.  
 CC -1- SIMILARITY: Belongs to the sulfotransferase family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC

```

CC or send an email to license@sib.ch).
CC -----
CC EMBL; M54942; AAA30679.1; -
CC DR EMBL; X56395; CAA39806.1; -
CC DR PIR; S29045; S29045.
CC DR HSP; P49891; IAQU.
CC DR InterPro; IPR00863; Sulfotransferase.
CC DR Pfam; PF00685; Sulfotransferase; 1.
CC DR ProDom; PD001218; Sulfotransferase; 1.
CC KW Transferase; Phosphorylation; Steroid-binding.
CC FT MOD_RES 216 216 PHOSPHORYLATION (BY PKA) (POTENTIAL).
CC FT MOD_RES 228 228 PHOSPHORYLATION (BY PKA) (POTENTIAL).
CC FT CONFLICT 117 117 S -> E (IN REF. 2).
CC FT CONFLICT 171 171 S -> E (IN REF. 2).
CC FT CONFLICT 248 248 M -> Q (IN REF. 2).
CC SQ SEQUENCE 295 AA; 34640 MW; 69147C73146913FD CRC64;

Query Match 15.0%; Score 309; DB 1; Length 295;
Best Local Similarity 31.0%; Pred. No. 6.3e-15;
Matches 95; Conservative 50; Mismatches 113; Indels 48; Gaps 14;

QY 49 SLPSP-NDYWGDT--NRLYKGFQWGYLVPGIMAFEDNFKARETDIILTLPLKAGTTWK 105
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 55
3 SLPSPSDYFGKLGIPMTKIEGFHN-----EFKAPDPLVIVTYPKSGTTWLS

QY 106 ALTFALTR-DVNHESPPTPLFFPHSCVQNLLEYLYMGRENTMPDLDMLNE--SPRLF 162
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 14;
56 EIIICMIYNGDVEKCKED---VIFN-----RVPLECSTEHVYKGVKQLNENASPRIV 105

QY 163 AGHIPVSLPASVLKSGTKIINISNRKSTFVSFWKFGNLI-----NPKLLD-LEKVD- 216
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 165
106 KSLPLVKLLPVSGFWKCKNKLIIYLSRNKADVVVSYFLIMVTAPDPDSFODFVEKFMG

QY 217 --IFASGISFCGPEWFOAFTNAASTNSNLLLSYEEMLEKPVENVKLAEPFGCGFTD 274
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 217
166 EYVGSWFPEHTKSWE-----KSKNPQVLFLFYEDMKENIRKVMKLEFLGRASD 217

QY 275 DEEKQIVDEIVKLSFDMKLNQVKNKSSYNSKIDNKH---FPRKEVDRWANYLTSE 331
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 271
218 E-----LVDKIIKHTSFQEMKNKP-STNYITLDFEVMNQKVSPPFMRKGDVGDKNHFVA

QY 332 MINKLE 337
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 272 LNEKFD 277

RESULT 7
SIC2_HUMAN STANDARD; PRT; 302 AA.
AC 075837;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Sulfotransferase 1C2 (EC 2.8.2.-) (SULT1C) (SULT1C#2).
GN SULT1C2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal lung;
RX MEDLINE=99069375; PubMed=9852044;
RA Sakakibara Y., Yanagisawa K., Katafuchi J., Ringer D.P., Takami Y.,
RA Nakayama T., Suiko M., Liu M.-C.;
RT "Molecular cloning, expression, and characterization of novel human
RT SULT1C sulfotransferases that catalyze the sulfonation of
RT N-hydroxy-2-acetylaminofluorene."
RL J. Biol. Chem. 273:33929-33935(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20247255; PubMed=10783263;
RA Freimuth R.R., Raftogiannis R.B., Wood T.C., Moon E., Kim U.-J., Xu J.,

```

```

RA Siciliano M.J., Weinshilboum R.M.;
RT "Human sulfotransferases SULT1C1 and SULT1C2: cDNA characterization,
RT gene cloning, and chromosomal localization.";
RL Genomics 65:157-165(2000).
CC -!- FUNCTION: Catalyzes the sulfate conjugation of many drugs,
CC xenobiotic compounds, hormones, and neurotransmitters. May be
CC involved in the activation of carcinogenic hydroxylamines. Shows
CC activity towards p-nitrophenol and N-hydroxy-2-acetylamine-
CC fluorene (N-OH-2AAF).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- TISSUE SPECIFICITY: Expressed at high levels in fetal lung and
CC kidney and at low levels in fetal heart, adult kidney, ovary and
CC spinal chord.
CC -!- SIMILARITY: Belongs to the sulfotransferase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).

CC EMBL; AF055584; AAC95519.1; -
CC DR EMBL; AF186263; AAF72810.1; -
CC DR HSSP; P50224; 1CJM.
CC Genew; HGNC:11457; SULT1C2.
CC GO; GO:0008146; P:sulfotransferase activity; NAS.
CC InterPro; IPR000863; Sulfotransferase.
CC Pfam; PF00685; Sulfotransferase; 1.
CC ProDom; PD001218; Sulfotransferase; 1.
KW Transferase.
FT CONFLICT 5
SQ SEQUENCE 302 AA; 35534 MW; DD2F486592E3B56D CRC64;

Query Match 14.8%; Score 304; DB 1; Length 302;
Best Local Similarity 30.0%; Pred. No. 1.5e-14;
Matches 93; Conservative 49; Mismatches 114; Indels 54; Gaps 13;

QY 59 DTMRLYKGFQWGYLVPGIMAFEDNFKARETDIILTLPLKAGTTWKALTFALTR-DVN 117
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 75
30 DTCDIWDKIW-----NFOAKPDDLIISTYPRAGTTWTQEIIVELIQNEGQVE 75

QY 118 HPS-SPTH---PLLFFPHSCVQNLLEYLYMGRENTMPDLDMLNESPRLFAGHIPYSLLPA 173
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 123
76 KSKRAPTHQRFPLEMKIPLSGGLE-----QAEAMP-----SPRLKTHLFFHLLPP 123

QY 174 SVLKSGTKIINISNRKSTFVSFWKPGNLINPDKLLD-----EKSDIFASGISFCGPEW 229
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 179
124 SLLEKNCKIIYVARNPKDNVSYTHFORM--NKALPAPGTWBEYFFETFLAG-KVCWGSW 179

QY 230 NFOASFTNAASTNSNLLLSYEEMLEKPVENVKLAEPFGCGFTDDEKQIVDEIVKLC 269
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 234
180 HEHVKGWWEAKDKRILYLFYEDMKQPKHEIOKLAEFIQKLLDK-----VLDKIVHYT 234

QY 290 SPDNINKQVKNKSSYNSKIDN---KHFFRKGEVRDWANYLTSEMTKKLETAGKINESE 346
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 287
235 SFDVWK-QNPWANYSSIPAEIMDHSISPPFMRKGAVGDKKKHFTVAQNERFD-----EDY 287

QY 347 KHLISNKNLT 356
Db |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||: 288 KKKMTDRLT 297

RESULT 8
SUAC RAT
ID SUAC RAT STANDARD; PRT; 304 AA.
AC P50237;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE N-hydroxyaryamine sulfotransferase (EC 2.8.2.-) (HAST-I).
GN ST1C1.

```



```

CC  EMBL; U08098; AAA82125.1; -
DR  EMBL; U20521; AAC50286.1; -
DR  EMBL; U20515; AAC50286.1; JOINED.
DR  EMBL; U20517; AAC50286.1; JOINED.
DR  EMBL; U20516; AAC50286.1; JOINED.
DR  EMBL; U20518; AAC50286.1; JOINED.
DR  EMBL; U20519; AAC50286.1; JOINED.
DR  EMBL; U20520; AAC50286.1; JOINED.
DR  EMBL; S77383; AAB34601.1; -
DR  EMBL; Y11195; CAA72079.1; -
DR  EMBL; U55764; AAB51658.1; -
DR  PIR; JC2229; JC2229
DR  PIR; LHY3; 19-JUN-02.
DR  Genew; HGNC:11377; STE.
DR  MIM; 60043; -
DR  GO; GO:0004304; F:estrogen sulfotransferase activity; TAS.
DR  GO; GO:0008202; P:steroid metabolism; TAS.
DR  InterPro; IPR000863; Sulfotransferase.
DR  Pfam; PF00685; Sulfotransferase; 1.
DR  ProDom; PD001218; Sulfotransferase; 1.
KW  Transferase; Steroid-binding; 3D-structure.
FT  BINDING 258 264 PAPS BINDING SITE (POTENTIAL).
SQ  SEQUENCE 294 AA; 35126 MW; 9EC923D20757D57 CRC64;

Query Match 12.4%; Score 255; DB 1; Length 294;
Best Local Similarity 27.4%; Pred. No. 4.3e-11;
Matches 85; Conservative 51; Mismatches 120; Indels 54; Gaps 13;

QY 41 YKRYREIVSSLPNDYNGDWTMLYKGFQWQGYLVPGIMAFEDNFKARETDIILTLTP 97
D 8 YKRFEEVHGIL-----MYKDF-----VKYWDNVEAFQAPDPLVIATPY 46
D 98 KAGTWTYKALTFAILTR-DVNHPSSTPHLLFFNPHSCVQNLLEYLYMGRENTMPDLMDLN 156
D 47 KSGTWTYSEIVLYMYKEGDEKED-----VFN-----RIFLECKRENLMNGYKQLD 96
QY 157 E--SPRLFAGHIPYSLLPASVLSGKTGIINISNRKSTFTVSWFKPGNLI-----NPKQLD 210
D 97 EBNSPRIKTHLPPELLPASFWKEDCKIILYLCRNADKQVAVSFYFFLMVAGHPNPGSFPE 156
QY 211 LEKSDVDFASGISCGFENWFOAEFTNAASTNSNLLLSYEEMLEKPEVNVKLAEFMGC 270
D 157 F---VEKFMGQVPGY-SWYKHVSWWEKSKRVLFLFYEDDKDIRKEVILKLIHFL-- 210
QY 271 GFTDDEKQGIIVDEIVKLCDFNLKQVNVKNGSSYNSKIDNRK---PFRKGEVRDWAY 327
D 211 ---ERKPSBELVDRIIHTSFQEMKNP-STNYTTLDPDEIMNQKLSPFMRKGITGDWKNH 266
QY 328 LTSEMIKLE 337
D 267 FTVALNEKFD 276

RESULT 10
SU03_RAT
ID SU03_RAT STANDARD; PRT; 295 AA.
AC P49889;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Estrogen sulfotransferase, isoform 3 (EC 2.8.2.4) (EST-3)
DE (Sulfotransferase, estrogen-prefering) (Estrogen sulfotransferase).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN (1)
RP SEQUENCE FROM N.A., AND SEQUENCE OF 258-265.
RC TISSUE=Liver;
RX MEDLINE=92261615; PubMed=1374939;
RA Demyan W.F., Song C.S., Kim D.S., Her S., Gallwitz W., Rao T.R.,
RA Slomczynska M., Chatterjee B., Roy A.K.;

```

```

RT "Estrogen sulfotransferase of the rat liver: complementary DNA
RL cloning and age- and sex-specific regulation of messenger RNA.";
RN Mol. Endocrinol. 6:589-597(1992).
RP (2)
RC SEQUENCE FROM N.A.
RX TISSUE=Liver;
RX MEDLINE=95161323; PubMed=7857871;
RA Falany J.L., Krasnykh V., Mikheeva G., Falany C.N.;
RT "Isolation and expression of an isoform of rat estrogen
RT sulfotransferase.";
RT J. Steroid Biochem. Mol. Biol. 52:35-44(1995)
CC -!- FUNCTION: Sulfation of estrone and estradiol. May control the
CC level of the estrogen receptor by sulfating free estradiol.
CC -!- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + estrone =
CC adenosine 3',5'-bisphosphate + estrone 3-sulfate.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- TISSUE SPECIFICITY: LIVER OF YOUNG MATURE MALES AND UTERUS
CC -!- DEVELOPMENTAL STAGE: Expressed only in the liver of young adult
CC animals (100 days old) and is absent in the prepubertal male (27
CC days old), senescent male (800 days old) and female liver.
CC -!- INDUCTION: Induced by androgens and suppressed by estrogens. The
CC expression is under the influence of pituitary growth hormone and
CC thyroid hormone. Is regulated by progesterone in the uterus.
CC -!- SIMILARITY: Belongs to the sulfotransferase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M86758; AAA41128.1; -
DR EMBL; S76489; AAB33441.1; -
DR PIR; A41930; A41930.
DR HSP; P49891; IAOU.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase; 1.
DR ProDom; PD001218; Sulfotransferase; 1.
KW Transferase; Steroid-binding; Multigene family.
FT BINDING 239 265 PAPS BINDING SITE (POTENTIAL).
FT CONFLICT 150 158 P -> Q (IN REF. 2).
FT CONFLICT 238 238 T -> I (IN REF. 2).
FT CONFLICT 235 295 L -> P (IN REF. 2).
SQ SEQUENCE 295 AA; 35415 MW; A77807A21DD2E7EB CRC64;

Query Match 12.3%; Score 253; DB 1; Length 295;
Best Local Similarity 26.7%; Pred. No. 5.9e-11;
Matches 85; Conservative 41; Mismatches 112; Indels 80; Gaps 12;

QY 40 MYKRYREIVSSLPNDYNGDWTMLYKGFQWQGYLVPGIMAFEDNFKARETDIILTLTPKA 99
D 20 MDRKFTK-----YWDV-----ETFLAPDDLLIIVTPKS 49
QY 100 GTTWTYKALTFAILTR-DVNHPSSTPHLLFFNPHSCVQNLLEYLYMGRENTMPDL----- 153
D 50 GSTWISSEIVDMYKEDV-----EKCKEDALF-----NRIPDLECRNED 88
QY 154 MLN-----ESPRLFAGHIPYSLLPASVLSGKTGIINISNRKSTFTVSWFKPGNLI-- 203
D 89 LINGIKQLKEKSPRIKTHLPPELLPASFWKEDCKIILYLCRNADKQVAVSFYFFLMVAGHPN 148
QY 204 --NPKQLDLEKSDVDFASGISCGFENWFOAEFTNAASTNSNLLLSYEEMLEKPEVNV 261
D 149 YNPKPSFSF---VEKFMGQVPGY-SWYKHVSWWEKSKRVLFLFYEDDKDIRREV 204
QY 262 KKLAEFMGCGFTDDEKQGIIVDEIVKLCDFNLKQVNVKNGSSYNSKIDNK--HFFKRG 319
D 205 VKLIEFL-----ERDPSAELVDRIIHTSFQEMKNPCTNYSMLEPMETMIDLVKVSFMRKG 259
QY 320 EYRDWANYLTSEMIKLE 337

```

```
Db 260 IVGDWKNHFPALRERPE 277
RESULT 11
SUO2_RAT
ID SUO1_RAT STANDARD; PRT; 295 AA.
AC P52844;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Estrogen sulfotransferase, isoform 1 (EC 2.8.2.4) (EST-1)
DE (Sulfotransferase, estrogen-prefering) (Estrone sulfotransferase).
GN STEL OR SUL1E1
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=Fischer 344; TISSUE=Liver;
RX MEDLINE=96305357; PubMed=6688469;
RA Rikke B.A.; Roy A.K.;
RT "Structural relationships among members of the mammalian
sulfotransferase gene family.";
RL Biochim. Biophys. Acta 1307:331-338 (1996).
CC -!- FUNCTION: Sulfation of estrone and estradiol. May control the
level of the estrogen receptor by sulfonylating free estradiol (By
similarity).
CC -!- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + estrone =
adenosine 3',5'-bisphosphate + estrone 3-sulfate.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the sulfotransferase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; U50204; AAB07680.1; -.
CC HSPF; P49891; IAAU.
CC InterPro; IPR000863; Sulfotransferase.
CC Pfam; PF00685; Sulfotransfer; 1.
CC ProDom; PD001218; Sulfotransferase; 1.
CC Transferrase; Steroid-binding.
FT BINDING 259 265 PAPS BINDING SITE (POTENTIAL).
SQ SEQUENCE 295 AA; 35509 MW; 696A12FDA92A12E CRC64;

Query Match 12.2%; Score 252; DB 1; Length 295;
Best Local Similarity 26.7%; Pred. No. 7e-11;
Matches 85; Conservative 41; Mismatches 112; Indels 80; Gaps 12;

QY 40 MYKREIVSLPSNDYGDITMLYKGFQMGVLPVGMAFEDNFKARETDILTLPKA 99
Db 20 MDKRRTK-----YWDI-----ETFLARPDLLIVTPKS 49
QY 100 GTTWTKALTFAILTR-DVNHPSPTPLFFNPHSCVQNLVYMGRENTMPDL-----D 153
Db 50 GSTWISIVDMYIEKGDV-----EKCKEDALF-----NRIPDECRNE 88
QY 154 MLN-----ESPRLPAGHYPYLLPASVLKSGTKIINSRNKRSTFVSFWKFGNLI-- 203
Db 89 LINGIKQKEREKSPRIVTHLPKLLPASFWKKNCKIILYCRNAKDVVVSYYFFFLIMKS 148
QY 204 --NPKLLDLEKSDVIFASGISFCGPEWNFQAFETNAASTNSNLLLSYEMLKPEVNY 251
Db 149 YENPKSFSEF---VEKMEGQVPYG-SWYDHVKSWEKSKNSRVLFMFYEDMKEDIRREY 204
QY 262 KKLAEFMCGGFTDDSEKQGVDEIVKLSFDNLKNQVKNNGSKINDK--HFRKRG 319
```

```
Db 205 VKLIEFL-----ERDPSAEIVDRIIQHTSFQEMKNPNCTNYSMLPETMIDLVSPFMRKG 259
QY 320 EYRDWANYLTSEMIKKLE 337
Db 260 IVGDWKNHFPALRERPE 277

RESULT 12
SUO2_RAT
ID SUO2_RAT STANDARD; PRT; 295 AA.
AC P52845;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Estrogen sulfotransferase, isoform 2 (EC 2.8.2.4) (EST-2)
DE (Sulfotransferase, estrogen-prefering) (Estrone sulfotransferase).
GN STE2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=Fischer 344; TISSUE=Liver;
RX MEDLINE=96305357; PubMed=6688469;
RA Rikke B.A.; Roy A.K.;
RT "Structural relationships among members of the mammalian
sulfotransferase gene family.";
RL Biochim. Biophys. Acta 1307:331-338 (1996).
CC -!- FUNCTION: Sulfation of estrone and estradiol. May control the
level of the estrogen receptor by sulfonylating free estradiol (By
similarity).
CC -!- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + estrone =
adenosine 3',5'-bisphosphate + estrone 3-sulfate.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the sulfotransferase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; U50205; AAB07681.1; -.
CC HSPF; P49891; IAAU.
CC InterPro; IPR000863; Sulfotransferase.
CC Pfam; PF00685; Sulfotransfer; 1.
CC ProDom; PD001218; Sulfotransferase; 1.
CC Transferrase; Steroid-binding.
FT BINDING 259 265 PAPS BINDING SITE (POTENTIAL).
SQ SEQUENCE 295 AA; 35364 MW; 149B5C9D46039AAF CRC64;

Query Match 12.2%; Score 252; DB 1; Length 295;
Best Local Similarity 26.0%; Pred. No. 7e-11;
Matches 85; Conservative 44; Mismatches 131; Indels 44; Gaps 12;

QY 49 SSLSF-----NDYWGDTM-RLYKGFQMGVLPVGMAFEDNFKARETDILTLPKAGTT 102
Db 3 TSMPEYVEVFGDFHGLMDKLFYKWE-----DVETTSARPDLLVIVTPKSGST 52
QY 103 WTKALTFAILTR-DVNHPSPTPLFFNPHSCVQNLVYMGRENTMPDLML--NESP 159
Db 53 WIGEVIMYIEKGDVEKCKEDA----IFN-----RIPYLECRNEDELINGIKQLEKESP 102
QY 160 RLFAGHYPYLLPASVLKSGTKIINSRNKRSTFVSFWKFGNLI-----NPKLLDLEKSV 215
Db 103 RIVKTHLPKLLPASFWKKNCKIILYCRNAKDVVVSYYFFFLIMKSYNPKSFSEF---V 159
QY 216 DIFASGISFCGPEWNFQAFETNAASTNSNLLLSYEMLKPEVNYVKLAEFMGCGFTDD 275
```

```
Db 160 EKFMEGQVPYG-SWYDHVKSWKSKNSRVLPFYEDMKEDIRREVVKLIEFL-----ER 213
Qy 276 BEKQIGVDEIVKLSFNLKQVKNKNGSSYNKIDNK--HFFRKGVRDWANYLTSEMI 333
Db 214 DPAELVDRIIQTSTFOEMKNNPCTNYSMLPETMIDLKVSFPMRKGIVGDWRNHFPEARL 273
Qy 334 KLE 337
Db 274 ERPE 277

RESULT 13
SU06_RAT STANDARD; PRT; 295 AA.
AC P49890;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Estrogen sulfotransferase, isoform 6 (EC 2.8.2.4) (EST-6)
DE (Sulfotransferase, estrogen-prefering) (Estrore sulfotransferase).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=95161323; PubMed=7857871;
RA Falany J.L., Kraenky V., Mikheeva G., Falany C.N.;
RT "Isolation and expression of an isoform of rat estrogen sulfotransferase.";
RL J. Steroid Biochem. Mol. Biol. 52:35-44 (1995).
CC -!- FUNCTION: Sulfation of estrone and estradiol. May control the level of the estrogen receptor by sulfurylating free estradiol.
CC -!- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + estrone = adenosine 3',5'-bisphosphate + estrone 3-sulfate.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- DEVELOPMENTAL STAGE: Expressed in males rats, but not in females.
CC -!- INDUCTION: Induced by androgens and suppressed by estrogens.
CC Expression is under the influence of pituitary growth hormone and thyroid hormone.
CC -!- SIMILARITY: Belongs to the sulfotransferase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S76490; A833442.1; -
DR HSP; P49891; IAOU.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
DR ProDom; PD001218; Sulfotransferase; 1.
KW Transferase; Steroid-binding; Multigene family.
FT BINDING 259 AA; 35302 MW; 9FBD5861AFDC9B05 CRC64;
SQ SEQUENCE 295 AA; 35302 MW; 9FBD5861AFDC9B05 CRC64;

Query Match 12.1%; Score 250; DB 1; Length 295;
Best Local Similarity 28.0%; Pred. No. 9.7e-11;
Matches 85; Conservative 45; Mismatches 130; Indels 44; Gaps 12;

Qy 49 SSIPS-----NDYWGDTM-RLYKGFQWQGYLVPGIMAFEDNFKARETDIILTLTKAGTT 102
Db 3 TSMPEYEVGDFHGVLDKLFKWE-----DVEFSARPDDLVLVTPKSGST 52
Qy 103 WKALTFAILTR-DVNHPSSTPHLLFFNPHSCVQNLEYLYMGRENTWPDLDML--NESP 159
Db 53 WIGEIVDMYKSGDVEKCKEDA----IFN-----RIPYLECRNEDLINGIKLKEKESP 102
```

```
Qy 160 RLFAGHIPYSLLPASVLKSGTKIINISRNKRKSTFVSFWKFGNLI-----NPKLLDLEKSV 215
Db 103 RIVKTHLPKALPASFWKCKNCKIIVLCRNADVVVYVYFLLIKSYPNPKSFSEF---V 159
Qy 216 DIPASGISPCGPEWNFQAEFTNAASTNSNLLLSVEEMLEKPVENVKLAEFMCGGFTDD 275
Db 160 EKFMEGQVPYG-SWYDHVKSWKSKNSRVLPFYEDMKEDIRREVVKLIEFL-----ER 213
Qy 276 BEKQIGVDEIVKLSFNLKQVKNKNGSSYNKIDNK--HFFRKGVRDWANYLTSEMI 333
Db 214 DPAELVDRIIQTSTFOEMKNNPCTNYSMLPETMIDLKVSFPMRKGIVGDWRNHFPEARL 273
Qy 334 KLE 337
Db 274 ERPE 277

RESULT 14
SUAR_MOUSE STANDARD; PRT; 291 AA.
AC P52840;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Aryl sulfotransferase (EC 2.8.2.1) (Phenol sulfotransferase) (PST-1)
DE (Sulfokinase) (Phenol/aryl sulfotransferase) (ST1A4).
GN SULTR1A OR ST1A1 OR STP1 OR STP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Liver;
RX MEDLINE=93144346; PubMed=8424956;
RA Kong A.-N.T., Ma M., Tao D., Yang L.;
RT "Molecular cloning of cDNA encoding the phenol/aryl form of sulfotransferase (mStp1) from mouse liver.";
RL Biochim. Biophys. Acta 1171:315-318 (1993).
CC -!- FUNCTION: Catalyzes O-sulfation of phenols, N-O-sulfation of minoxidil and tyrosine esters (By similarity).
CC -!- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + a phenol = adenosine 3',5'-bisphosphate + an aryl sulfate.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the sulfotransferase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L02331; -; NOT_ANNOTATED_CDS.
DR HSP; P50224; ICMJ.
DR MGD; MGI:102896; Sult1a1.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
DR ProDom; PD001218; Sulfotransferase; 1.
KW Transferase; Steroid metabolism.
FT BINDING 255 AA; 261 MW; AD07187DFAS1D96 CRC64;
SQ SEQUENCE 291 AA; 33974 MW; AD07187DFAS1D96 CRC64;

Query Match 12.1%; Score 248.5; DB 1; Length 291;
Best Local Similarity 27.6%; Pred. No. 1.2e-10;
Matches 81; Conservative 42; Mismatches 121; Indels 49; Gaps 10;

Qy 48 VSSLPENDYWGDTMRLYKGFQWQGYLVPGIMAFEDNFKARETDIILTLTKAGTTWKAL 107
Db 11 VKGIPLIKYPAETMEQLQ-----NFTAWPDDVLISITPKSGTNNWSEI 53
```





Db 91 NGIKQLKESEPRIVKTHLPKLLPASFWEKNCMIYLCRNAKVAVSVYYFLMLITSYP 150  
QY 204 NPKLLDLEKSYDIPASGISFCGPEWNPQAEFTNAASTNSNLLLSYEEMLEKPVENVK 263  
Db 151 NPKSFSEF---VEKFMQGVPIG-SWYDHVKAWWEKSKNSRYLFMFYEDMKEDIRREYVK 206  
QY 264 LAEFMCGGFTDBEKQGIYDEIVKLCSDNLKNOQVKNKNGSSY-----NSKIDNKH 315  
Db 207 LIEFL-----ERKPSAELVDRIIQTSTFOEMKNP-----STNYTWMPEEMNQV--SPF 255  
QY 316 FRKGEVRDWANYLTSEMICKLE 337  
Db 256 MRKGIIGDWKNHFPALRERFD 277

Search completed: March 20, 2004, 01:38:51  
Job time : 12 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2004, 01:34:18 ; Search time 35 Seconds

(without alignments)  
3524.789 Million cell updates/sec

Title: US-09-854-122-16

Perfect score: 2059

Sequence: 1 TRGITGIACVACPLINAGIL.....CLLSYSFVNNFILLKKKK 391

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL 25:

- 1: sp\_archaea:
- 2: sp\_bacteria:
- 3: sp\_fungi:
- 4: sp\_human:
- 5: sp\_invertebrate:
- 6: sp\_mammal:
- 7: sp\_mhc:
- 8: sp\_organelle:
- 9: sp\_phase:
- 10: sp\_plant:
- 11: sp\_rodent:
- 12: sp\_virus:
- 13: sp Vertebrate:
- 14: sp Unclassified:
- 15: sp\_virus:
- 16: sp\_bacteriap:
- 17: sp\_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	616	29.9	329	Q9M1V1	Q9M1V1 arabidopsis
2	602	29.2	323	Q9M1V2	Q9M1V2 arabidopsis
3	578.5	28.1	331	Q9FX56	Q9FX56 arabidopsis
4	572.5	27.8	331	Q8LEH6	Q8LEH6 arabidopsis
5	567	27.5	346	Q9FZ80	Q9FZ80 arabidopsis
6	562.5	27.3	338	Q9C9D0	Q9C9D0 arabidopsis
7	561.5	27.3	338	Q8LB09	Q8LB09 arabidopsis
8	558.5	27.1	350	Q8LA18	Q8LA18 arabidopsis
9	558.5	27.1	350	Q9C9C9	Q9C9C9 arabidopsis
10	555.5	27.0	359	Q8LSA7	Q8LSA7 arabidopsis
11	551.5	26.8	359	Q9FLA5	Q9FLA5 arabidopsis
12	549.5	26.7	359	Q940P4	Q940P4 arabidopsis
13	548.5	26.6	351	Q8RV79	Q8RV79 arabidopsis
14	546	26.5	325	Q82410	Q82410 brassica na
15	545	26.5	331	Q9ZPQ6	Q9ZPQ6 arabidopsis
16	544.5	26.4	326	Q9FZ91	Q9FZ91 arabidopsis

17	536.5	26.1	351	10	Q9FX55	Q9FX55 arabidopsis
18	533	25.9	334	10	Q8LMK8	Q8LMK8 oryza sativ
19	529.5	25.7	323	10	Q82408	Q82408 brassica na
20	528.5	25.6	332	10	Q84270	Q84270 oryza sativ
21	527	25.6	324	10	Q9ZPQ5	Q9ZPQ5 arabidopsis
22	525.5	25.5	333	10	Q82330	Q82330 arabidopsis
23	524	25.4	358	10	Q7XMT7	Q7XMT7 oryza sativ
24	517.5	25.1	346	10	Q7XKJ3	Q7XKJ3 oryza sativ
25	510.5	24.8	340	10	Q9FLA6	Q9FLA6 arabidopsis
26	510.5	24.8	347	10	Q8GZ53	Q8GZ53 arabidopsis
27	500.5	24.3	324	10	Q82409	Q82409 brassica na
28	499	24.2	331	10	Q9FG94	Q9FG94 arabidopsis
29	486.5	23.6	314	10	Q9STQ6	Q9STQ6 arabidopsis
30	446.5	21.7	346	10	Q8L452	Q8L452 oryza sativ
31	407	19.8	273	10	Q8RUC1	Q8RUC1 arabidopsis
32	403.5	19.6	287	10	Q84269	Q84269 oryza sativ
33	356	17.3	250	10	Q84NE0	Q84NE0 tritium mo
34	303	14.7	295	6	Q9SMF8	Q9SMF8 sus scrofa
35	278.5	13.5	301	6	Q46640	Q46640 oryctolagus
36	273.5	13.3	304	13	Q7ZYH0	Q7ZYH0 xenopus lae
37	273	13.3	299	13	Q7T1C8	Q7T1C8 brachydanio
38	270	13.1	316	11	Q8C301	Q8C301 mus musculus
39	269	13.1	301	13	Q7T1C7	Q7T1C7 brachydanio
40	268.5	13.0	301	13	Q7T2V2	Q7T2V2 brachydanio
41	266.5	12.9	307	13	Q90WR6	Q90WR6 gallus gall
42	264.5	12.8	294	13	Q7S293	Q7S293 xenopus lae
43	262	12.7	301	13	Q7ZUS4	Q7ZUS4 brachydanio
44	261	12.7	291	11	Q91X36	Q91X36 mus musculus
45	259	12.6	294	4	Q8N6X5	Q8N6X5 homo sapien

#### ALIGNMENTS

#### RESULT 1

Q9M1V1 ID Q9M1V1 PRELIMINARY; PRT; 329 AA.

AC Q9M1V1  
DT 01-OCT-2000 (TRENBLrel. 15, Created)  
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
DE Sulfotransferase-like protein.  
T14D3.20.  
GN Arabidopsis thaliana (Mouse-ear cross).  
OS Arabidopsis thaliana (Mouse-ear cross).  
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC euroids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Jordan N., Bangert S., Wiedemann R., Voss H., Unseld M., Mewes H.W.,  
RA Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.,  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA EU Arabidopsis sequencing project;  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL138649; CAB72146.1; -  
DR PIR; T47448; T47448.  
DR HSSP; P50224; 1CTM.  
DR GO; GO:0008146; F:sulfotransferase activity; IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.  
DR InterPro; IPR000863; Sulfotransferase.  
DR Pfam; PF00685; Sulfotransfer; 1.  
DR ProDom; PD001218; Sulfotransferase; 1.  
KW Transferase.  
SQ SEQUENCE 329 AA; 38069 MW; 9D04227A483A630C CRC64;

Query Match 29.9%; Score 616; DB 10; Length 329;  
Best Local Similarity 41.2%; Pred. NO. 1.2e-34;  
Matches 132; Conservative 62; Mismatches 100; Indels 26; Gaps 9;

QY 29 KNEQKEEDSKMYKRYRIVSSLPN-DYWGDTNRLYKGFQMGVLPVGINAFEDNFKAR 87

```

11 RNDLSEES-----KTLISLPSDKNSTGVNVCKYQCMWTPPILOGLVNFQKNFKPQ 63
88 ETDIITLTPKAGTTWKALTALITRDVNHPS-SPTHELLFFNPHSCVQNL-PLYMGR 145
64 DTDIIVASFPKGTWKLKALTALVRS-KHSHDDHLLSDNPAVLSPELMYLCS 122
146 ENTMPDMLNTPRIFAGHPYSLPASVLKSGTKINISNRKSTFVFWKF----- 199
123 EN--PDLTKFSSSRLLFTMPSHLTQELGKSTCKIVYMSRNVKDTLVSYWHFFCKQT 180
200 -GNLNPDKLLDLKSVDFASGISFCPEWFOAEFTNAASTNSN-LLLSYEMLKRP 257
181 DNILS-----SVEDTFEFCRGVNFPGFWHVLVSYRGSLEDNHLVFMKFEEMKEP 235
258 VENVKLAFFMGCGTDDDEKQIVDEIVKLCSFNLKQVKNKSGSYNSKIDNKHFR 317
236 REQIKRLAFLGCLFTKEEESGLVDEIIDLSRLNLSLEINKTKLHSTGRENKTFR 295
318 KGEVDRWANYLTSEMICKLE 337
296 KGEVGDKWNYLTPEMENKID 315

RESULT 2
Q9M1V2 PRELIMINARY; PRT; 323 AA.
AC Q9M1V2;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Sulfotransferase-like protein.
TI4D3.10.
GN Arabidopsis thaliana (Mouse-ear cress).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Jordan N., Bangert S., Wiedemann R., Voss H., Unseld M., Mewes H.W.,
RA Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; ALI38649; CAB72145.1; -
DR PIR; T47447; T47447.
DR HSP; P50224; ICGM.
DR GO; GO:000846; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase; 1.
DR ProDom; PD001218; Sulfotransferase; 1.
DR Transferase.
SQ SEQUENCE 323 AA; 37501 MW; 00F9259B9CCF6389 CRC64;

Query Match 29.2%; Score 602; DB 10; Length 323;
Best Local Similarity 42.6%; Pred. No. 1.1e-33;
Matches 127; Conservative 51; Mismatches 112; Indels 8; Gaps 6;

QY 45 REIVSSLPNS-DYWGDTMRLYKGFQMGVLPVPGIMAFEDNFKARETDIITLTPKAGTTW 103
DB 15 KTLISLPSDKNSTGVNVCKYQCMWTPPILOGLVNFQKNFKPQ 74
QY 104 TKALTFAITRDVNHPSPTLFFNPHSCVQNL-PLYMGRNTMPDMLNTPRIF 162
DB 75 LKALTFAITLHRSQPSHDDHLLSNPHVLPVFEIDLYLRSEN--PDLTKFSSSRLLF 132
QY 163 AGHIPYSLPASVLKSGTKINISNRKSTFVFWKFGNLIINPKD--LLDLKSVDFEAS 220
DB 133 STVPSHTLQBLKSGTCKIVYISRNVDITLVSYWHFFTKQTDEKIISFDFEFNFCR 192

```

```

221 GISFCPEWFOAEFTNAASTNSN-LLLSYEMLKRPVENVKLAFFMGCGTDDDEKQ 279
193 GVSIFGPFWDHVLVSYRGSLEDNHLVFMKFEEMKAEPRDQIKKFAEFLGCPFTKEEBS 252
280 GIYDEIVKLCSFNLKQVKNKSGSYNSKIDNKHFRKGEVDEWANYLTSEMICKLE 337
253 GSVDEIIDLSRLNLSLEINKTG-KUNSGRENKMFRRKGEVGDWQNYLTPEMENKID 309

RESULT 3
Q9FX56 PRELIMINARY; PRT; 331 AA.
AC Q9FX56;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE T6J4.16 protein.
GN T6J4.16.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA Alcafi H., Nguyen M., Lam B., Southwick A., Miranda M., Brooks S.,
RA Buehler E., Chao Q., Chin C., Chou J., Choi E., Gonzalez A.,
RA Hwang B., Johnson-Hopson C., Khan S., Kim C., Koo T., Lee J.M.,
RA Lenz C., Liu A., Liu S., Mukharasy N., Pham P., Sakano H., Shinn P.,
RA Toriumi M., Vaysberg M., Yu G., Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC011810; AAG09547.1; -
DR HSP; P50224; ICGM.
DR GO; GO:000846; F:sulfotransferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase; 1.
DR ProDom; PD001218; Sulfotransferase; 1.
SQ SEQUENCE 331 AA; 37709 MW; 45DC8565C780E324 CRC64;

Query Match 28.1%; Score 578.5; DB 10; Length 331;
Best Local Similarity 40.2%; Pred. No. 4.8e-32;
Matches 125; Conservative 60; Mismatches 117; Indels 9; Gaps 7;

QY 31 EQEKEEDSKMYKYREIVSSLPNS-DYWGDTMRLYKGFQMGVLPVPGIMAFEDNFKARET 89
DB 12 EEEEEENQSEETKSLISLPSIDCSGKLYKQCYDKDILQAILNFKNFQPOET 71
QY 90 DIILTLPKAGTTWKALTALITRDVNHPSPTLFFNPHSCVQNL-PLYMGRNT 148
DB 72 DIIVASFPKGTWKLKALTALVRSKH--TSDNHPLLTNPHLVPYLEDLYL--KSS 127
QY 149 MPDMLNTPRIFAGHPYSLPASVLKSGTKINISNRKSTFVFWKFGNLIINPKD 207
DB 128 KPDLTLPSSSPRLFTSMGFDALKVPKESPKIVVVCNVKDVLSLWCFENSMGSEN 187
QY 208 LLDLEKSVDFASGISFCPEW-NQAEFTNAASTNSNILLISYEMLKRPVENVKLA 266
DB 188 NLSLEALFESLCSGVNLCGLWENLVGYWRGSLEDNHLVFLYRLEELKTPRVOIKLAE 247
QY 267 FMWCGGTDDDEKQIVDEIVKLCSFNLKQVKNKSGSYNSKIDNKHFRKGEVDRW 326
DB 248 FLDGPFKEEDSGVDKILELCSRLNLSLEINKTG-SLSEGVSPKSPFRKGEVGDWKS 306
QY 327 YLTSEMICKLE 337
DB 307 YMTPEMENKID 317

RESULT 4
Q9LEH6

```

```

ID QBLEH6 PRELIMINARY; PRT; 331 AA.
AC QBLEH6;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Steroid sulfotransferase, putative.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Haas B.J., Volfovsky N., Town C.D., Trukhan M., Alexandrov N.,
RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RA "Full-length messenger RNA sequences greatly improve genome
RT annotation.";
RL Genome Biol. 0:0-0(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Brover V., Trukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA Feldmann K.;
RA "Full-length cDNA from Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY085411; AM62638.1; -.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
DR ProDom; PD001218; Sulfotransferase; 1.
KW Transferase.
SQ SEQUENCE 331 AA; 37709 MW; 1AB5475049C64AE9 CRC64;

Query Match 27.8%; Score 572.5; DB 10; Length 331;
Best Local Similarity 39.5%; Pred. No. 1:3e-31;
Matches 123; Conservative 61; Mismatches 118; Indels 9; Gaps 7;

QY 31 EQBEKEDSKMYKRYREIVSSLPN-DYWGDTWELYKFGWGMGLVPGIMAFEDNFKARET 89
DB 12 EEEEEEENQSETEKLSLSDIDCSGTYLYKGYCWDYDILQALINFNKNQFQPT 71

QY 90 DIILTLPKAGTTWKALTFALTIRDVNHPSSPTPLFPNPHSCVQNYLEY-LYMGRENT 148
DB 72 DIIVASFPKSGTTWLKALTFALAQKSH--TSDNHPLLTHNPHLVPYLELDLYL--KSS 127

QY 149 MPDLML-NESPRLFAGHIPYSLLPASVLKSGTKIINISNRKSTFTVSPKFGNLIINPK 207
DB 128 KPLDTKLPSSPSLFSTHMSFDALKVPLKESPKIVIVCRNVNDVLISLWCFENSGEN 187

QY 208 LLBLEKSVDFASG13FCGPPEW-NFOAEFTNAASTNSNLLLSYEEMLEKPVENVKCLAE 266
DB 188 NLSLEALFESLCSGVNLGGLWENLVGWRGSLDPAKHLFLRYEELKTEPRVQIKRLAE 247

QY 267 FMCGGTDDDEKOGIYDEIVKLCSFONLKNQVKNKSSVNSKIDNKHFRKGEYRDWAN 326
DB 248 FLCQPFTEBEDSGVDKILELCSLNLGLSLEINKTG-SLSEGVSPKSPFRKGEVDWKS 306

QY 327 YLTSEMIKKLE 337
DB 307 YMTPEMENKID 317

RESULT 5
Q9FZ80 PRELIMINARY; PRT; 346 AA.
AC Q9FZ80;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative flavonol sulfotransferase (Hypothetical protein) (Putative
DE flavonol 4'-sulfotransferase).
GN F25116.7 OR AT1G18590/P25116.11 OR AT1G18590.
OS Arabidopsis thaliana (Mouse-ear cress).

```

```

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA Altafi H., Nguyen M., Lam B., Southwick A., Miranda M., Brooks S.,
RA Buehler E., Chao Q., Chin C., Chiou J., Choi E., Gonzalez A.,
RA Hong B., Johnson-Hopson C., Khan S., Kim C., Koo T., Lee J.M.,
RA Lenz C., Liu A., Liu S., Mukharsky N., Pham P., Sakano H., Shinn P.,
RA Toriumi M., Vaynsberg M., Yu G., Ecker J., Theologis A., Davis R.W.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,
RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,
RA Hayashizaki Y., Shinozaki K.;
RL "Arabidopsis thaliana full-length cDNA.";
RT Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,
RA Onodera C.S., Quach H.L., Tang C., Toriumi M., Wong C., Wu H.C.,
RA Yu G., Yuan S., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,
RA Ishida J., Jones T., Kamiya A., Kawai J., Kim C.J., Narusaka M.,
RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,
RA Southwick A., Tripp M.G., Wu T., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RL "Arabidopsis Open Reading Frame (ORF) Clones.";
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC026238; AAF98415.1; -.
DR EMBL; AK117408; BAC42075.1; -.
DR EMBL; BT005398; AAC63818.1; -.
DR PIR; E86319; E86319.
DR HSP; P50224; 1CJM.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
DR ProDom; PD001218; Sulfotransferase; 1.
KW Hypothetical protein; Transferase.
SQ SEQUENCE 346 AA; 39912 MW; 84208C67260FD6B8 CRC64;

Query Match 27.5%; Score 567; DB 10; Length 346;
Best Local Similarity 39.6%; Pred. No. 3.2e-32;
Matches 137; Conservative 56; Mismatches 123; Indels 30; Gaps 11;

QY 23 EKCFGKNEQKEBDSKMYKRYREIVSSLPNDYW--GDTMRLYKFGWGMGLVPGIMAF 80
DB 17 ELASSSPSEFKNQ-----KHVQIIATLPHKQGRPKDPFVEYGGHMLQLLEGLLHA 71

QY 81 EDNFKARETDIILTLPKAGTTWKALTFALTIR--DVNHPSSPTPLFPNPHSCVQN 137
DB 72 QKFRPNDPFVCSYPTGTGTWLKALTFALANKSFDVS-----TNPLKRNPHFVPY 126

QY 138 LEYLKMGRENTMPDLML-NESPRLFAGHIPYSLLPASVLKSGTKIINISNRKSTVSVF 196
DB 127 IEIDF---PFFPSVDVLKDBGNLTFSTHIPYDLLPESVVKSGCKIVIVWRDPKDTFVM 182

QY 197 WKFGNLIINPKD--LLBLEKSVDFASG13FCGPPEW-NFOAEFTNAASTN-SNLLLSSEEM 253
DB 183 WTFAPKERSQOGPVVSEAEAFDKYCOGLSAYGPLYDHLVGLWYKAYQANPOQILFLKYETM 242

QY 254 LEKPVENVKLAEEFGCGFTDDEEKQIGIVDEIVKLCSFDMLKNQVKNKSSVNSK---I 310
DB 243 RADPLPYKRLAEFGVGYFKEEBEGNVKVKLCSFETLKNLEANKGEKREDRPVY 302

QY 311 DNKHFFRKGGEYRDWANYLTSEMIKKLETAGKINESEK--HLLSNK 354
DB 303 ANSAYFRKGVGDWQNLTPENVARID--GLMEBKFGTGFLLSSKS 346

```

```

33 EKEDSKMYKRYREIVSSLPNDYW--GDTWRLYKGFQWQGYLVPGIMAFEDNFKARETD 90
14 ELTEFEKTQKQYQDFIATLPKSKGWRPDEILTYQGGHWQECLEGLFHAKDHFPEARPTD 73
91 IILTLTPKAGTTWTWKALTFAITRDVNHPSPTPLFFNPHSCVQNLVLYMGRENTMP 150
74 FLVCSYPTGTTWLTALTYAIVNR--SRYDDAANPLKRNPHFVPYVEIDFA---FYP 127
151 DLDMLNESPR-LFAGHIPYSLLPASVLKSGTKYINISRNKSTFVSFWKFGNLPDK-- 207
128 TVDLQDRKNPLFSTHPIENGLLPDSIVNSGCKMYIWRDPKDTFISWMTF---LHKERSQ 184
208 ---LLDLEKSDVIDFASGISFGCPENWFOAETNAASTN-SNLLLSYEEMLEKEVENYVK 263
185 EQQLASLEDSDFMCKGVSUVGYLDHVLGVYKAYQENPDRIILFLRYETMEANLPLFVKR 244
264 LAEFMGCFTDDEBKQGVDEIVKLCSPDNKQOVKNKNGSSYNK-----DNKHFFRKE 320
245 LAEFMGYGTDEEENGVAEKVVKLCSPETLNLEANKKREDRDPAYANSAYFRKKG 304
321 VRDMANYLTSEMICKLETAGKINSEKH--LLSNKN 354
305 VGDWANYLTPEMARID--GLVEKFKDTGLLQHDN 338

RESULT 7
O8LB09 PRELIMINARY; PRT; 338 AA.
ID O8LB09;
AC O8LB09;
DT 01-OCT-2002 (TREMELrel. 22, Created)
DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Putative flavonol sulfotransferase.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Haas B.J., Volfovsky N., Town C.D., Troughan M., Alexandrov N.,
RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RT "Full-length messenger RNA sequences greatly improve genome
RT annotation.";
RL Genome Biol. 0:0-0 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Brover V., Troughan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA Feldmann K.;
RT "Full-length cDNA from Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY087493; AA65036.1;
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
DR ProDom; PD001218; Sulfotransferase; 1.
KW Transferase.
SQ SEQUENCE 338 AA; 39233 MW; ED8A7E0E906866F3 CRC64;

Query Match 27.3%; Score 561.5; DB 10; Length 338;
Best Local Similarity 38.4%; Pred. No. 7.4e-31;
Matches 129; Conservative 59; Mismatches 123; Indels 25; Gaps 10;

OY 33 EKEDSKMYKRYREIVSSLPNDYW--GDTWRLYKGFQWQGYLVPGIMAFEDNFKARETD 90
14 ELTEFEKTQKQYQDFIATLPKSKGWRPDEILTYQGGHWQECLEGLFHAKDHFPEARPTD 73
91 IILTLTPKAGTTWTWKALTFAITRDVNHPSPTPLFFNPHSCVQNLVLYMGRENTMP 150
74 FLVCSYPTGTTWLTALTYAIVNR--SRYDDAANPLKRNPHFVPYVEIDFA---FYP 127
151 DLDMLNESPR-LFAGHIPYSLLPASVLKSGTKYINISRNKSTFVSFWKFGNLPDK-- 207
128 TVDLQDRKNPLFSTHPIENGLLPDSIVNSGCKMYIWRDPKDTFISWMTF---LHKERSQ 184
208 ---LLDLEKSDVIDFASGISFGCPENWFOAETNAASTN-SNLLLSYEEMLEKEVENYVK 263
185 EQQLASLEDSDFMCKGVSUVGYLDHVLGVYKAYQENPDRIILFLRYETMEANLPLFVKR 244
264 LAEFMGCFTDDEBKQGVDEIVKLCSPDNKQOVKNKNGSSYNK-----DNKHFFRKE 320
245 LAEFMGYGTDEEENGVAEKVVKLCSPETLNLEANKKREDRDPAYANSAYFRKKG 304
321 VRDMANYLTSEMICKLETAGKINSEKH--LLSNKN 354
305 VGDWANYLTPEMARID--GLVEKFKDTGLLQHDN 338

RESULT 7
O8LB09 PRELIMINARY; PRT; 338 AA.
ID O8LB09;
AC O8LB09;
DT 01-OCT-2002 (TREMELrel. 22, Created)
DT 01-OCT-2002 (TREMELrel. 22, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Putative flavonol sulfotransferase.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Haas B.J., Volfovsky N., Town C.D., Troughan M., Alexandrov N.,
RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RT "Full-length messenger RNA sequences greatly improve genome
RT annotation.";
RL Genome Biol. 0:0-0 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Brover V., Troughan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA Feldmann K.;
RT "Full-length cDNA from Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY087493; AA65036.1;
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
DR ProDom; PD001218; Sulfotransferase; 1.
KW Transferase.
SQ SEQUENCE 338 AA; 39233 MW; ED8A7E0E906866F3 CRC64;

Query Match 27.3%; Score 561.5; DB 10; Length 338;
Best Local Similarity 38.4%; Pred. No. 7.4e-31;
Matches 129; Conservative 59; Mismatches 123; Indels 25; Gaps 10;

```





KW Transferase.  
SQ SEQUENCE 354 AA; 40887 MW; E540A1449966A36C CRC64;  
Query Match 26.8%; Score 551.5; DB 10; Length 354;  
Best Local Similarity 36.2%; Pred. No. 3.8e-30;  
Matches 122; Conservative 69; Mismatches 136; Indels 11; Gaps 8;  
QY 7 IAVACLPLIMAGILALEKFCGSKNEQKEEDSKMYKREIVSVSLPNDYWGDT-MELYK 65  
DB 4 IPMAIPSFMSCHKLELLKGGTRDVPKAEDEGLSCFQEMDLSPKRGWRIRYLYLQ 63  
QY 66 GFWQGVILVPGIMAFENFKAETDIIITLTPKAGTTWTATPAILTRDNNHP--SSPT 123  
DB 64 GFWQCAKEIOAIMSFQKHFQSLNDVLTATPKSGTTLKALTFTILNRHFRDP 123  
QY 124 HPLTFSPHDLVPFFEYKLYANGD--VPDLSGL-ASPRTFATHLPFGSLKETEIKPKVKV 180  
DB 124 HPLTFSPHDLVPFFEYKLYANGD--VPDLSGL-ASPRTFATHLPFGSLKETEIKPKVKV 180  
QY 183 INISNRKSTFTSVFWKFGNLPDKL--LDLEKSVDFISGIFSGPFWNFQAEF-TNAA 239  
DB 181 VYLRCNPPDTFTSSWHYTNNIKSESVSPVLLDQAFDLYCRGVIGFPGFWEHMLGYWRESL 240  
QY 240 STNSNLLLSYEEMLEKPVENVKLAERMGCGFTDDSEKQGVDEIVKLCSFNLKNQOV 299  
DB 241 KRPEKVFRLRYEDLKDDIETNLKLTATFLFPFTESEERKGVYKAIACLSFENLKKLEV 300  
QY 300 NKGSSYNKSIDNKHFRKGEVRDWAYNLTSMTSEMIKKL 336  
DB 301 NKSNSIKN-PENRFLFRKGEVSDWVYLSPSQVERL 336  
RESULT 12  
Q940P4 PRELIMINARY; PRT; 359 AA.  
AC Q940P4;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE AT5G07010/MOJ9.19.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Shinn P., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C.,  
RA Shinn P., Jones T., Carninci P., Dale J.M., Goldsmith A.D.,  
RA Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,  
RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,  
RA Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,  
RA Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,  
RA Tang C.C., Toriumi M., Yamada K., Yamamura Y., Yu G., Yu S.,  
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;  
RT "Arabidopsis cdna clones."  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
DR GO: 0008146; F:sulfotransferase activity; IEA.  
DR EMBL; AY054219; AAL06879.1;  
DR InterPro; IPR000863; Sulfotransferase.  
DR Pfam; PF00685; Sulfotransferase.  
DR ProDom; PD001218; Sulfotransferase; 1.  
SQ SEQUENCE 359 AA; 41378 MW; 061A4F79873CF44 CRC64;  
Query Match 26.7%; Score 549.5; DB 10; Length 359;  
Best Local Similarity 35.6%; Pred. No. 5.4e-30;  
Matches 122; Conservative 69; Mismatches 141; Indels 11; Gaps 8;  
QY 1 TRGITGIAVACLPLIMAGILALEKFCGSKNEQKEEDSKMYKREIVSVSLPNDYWGDT 60  
DB 3 TSSMKSIPMAIPSFMSCHKLELLKGGTRDVPKAEDEGLSCFQEMDLSPKRGWRIR 62  
QY 61 -MRLYKGFQWQGVILVPGIMAFENFKAETDIIITLTPKAGTTWTATPAILTRDNNHP 119

DB 63 YLYLFGQFWQCAKEIOAIMSFQKHFQSLNDVLTATPKSGTTLKALTFTILNRHFRDP 122  
QY 120 --SSPTHLLFPNPHSCVQNLEY-LYMGRENTMPDLMLNSPRLFAGHIPYSLPASVL 176  
DB 123 VASSTNHPLFTSNPHDLVPFFEYKLYANGD--VPDLSGL-ASPRTFATHLPFGSLKETE 179  
QY 177 KSGTKIINISNRKSTFTSVFWKFGNLPDKL--LDLEKSVDFISGIFSGPFWNFQAE 234  
DB 180 KPGVKVYLYCRNPPDTFTSSWHYTNNIKSESVSPVLLDQAFDLYCRGVIGFPGFWEHMLG 239  
QY 235 F-TNAASTNSNLLLSYEEMLEKPVENVKLAERMGCGFTDDSEKQGVDEIVKLCSFND 293  
DB 240 YWRESLKPEKVFRLRYEDLKDDIETNLKLTATFLFPFTESEERKGVYKAIACLSFEN 299  
QY 294 LKNQGVNKGSSYNKSIDNKHFRKGEVRDWAYNLTSMTSEMIKKL 336  
DB 300 LKLELVNKSNIK-PENRFLFRKGEVSDWVYLSPSQVERL 341  
RESULT 13  
Q8RV79 PRELIMINARY; PRT; 351 AA.  
ID Q8RV79  
AC Q8RV79;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Putative steroid sulfotransferase.  
GN AT2G03750.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,  
RA Barnett M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.-I.,  
RA Carrara A.J., Creasy T.H., Buehl C.R., Town C.D., Nieman W.C.,  
RA Fraser C.M., Venter J.C.;  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,  
RA Palm C.J., Bowser L., Jones T., Ban J., Carninci P., Chen H.,  
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,  
RA Ecker J., Theologis A., Davis R.W.;  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Nguyen M., Karlin-Neumann G., Southwick A., Tripp M., Miranda M.,  
RA Palm C.J., Bowser L., Jones T., Ban J., Carninci P., Chen H.,  
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,  
RA Ecker J., Theologis A., Davis R.W.;  
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC005836; RAD20077.2;  
DR EMBL; AY092961; AAM12960.1;  
DR EMBL; BT002108; AAN72119.1;  
DR GO: 0008146; F:sulfotransferase activity; IEA.  
DR GO: 0008146; F:sulfotransferase activity; IEA.  
DR InterPro; IPR000863; Sulfotransferase.  
DR Pfam; PF00685; Sulfotransferase.  
DR ProDom; PD001218; Sulfotransferase; 1.  
KW Transferase.  
SQ SEQUENCE 351 AA; 40838 MW; 2B8B7F61060333FA CRC64;  
Query Match 26.6%; Score 548.5; DB 10; Length 351;  
Best Local Similarity 37.4%; Pred. No. 6.1e-30;



Matches 125; Conservative 61; Mismatches 125; Indels 23; Gaps 10;

QY 22 LEKFGS-----KNEQEKE-----EDSXYKRYREIVSSLPSPN-DYMGDTWRLYKGF 67

Db 8 LSLCFKSWEQIITWEASKEAHLNPNYKDDNVSQETKNLITSLPSDKDFMGYGLNYKGC 67

QY 68 WQMGVLPVGINAFEDNFKARETDILITLTPKAGTWTWKALTFALLTRDYNHPSPTPL 127

Db 68 WYFNTLQAVLDVQKHPRDITLALSLPKGTTWLKSLFPAVVRKRGRTQTPTPL 127

QY 128 FPNHSCVQNLLEY-LYMGRENTMPDMLNESPRLFAGHIPYSLLPASVLSKSGTKIINIS 186

Db 128 LQNPEDLVFLEVELYANSQ--IPDLAKYS-SPMIFSTHMLQALREATTKA-CKTVVVC 183

QY 187 NNRKSTFVSFWKFGNLINPKD--LDLEKSVDDIPASGISFCGPEWFOAEF-TWAASNS 243

Db 184 RGIKDTFVSGHRYNMLHRTMQDQATFELMFDAYCGVLLYGPYWEHVLVSNYKGSLEAKE 243

QY 244 NLLLSYREMLEKPVENKVKLAERFCGFTDDEKQGVDBIVKLCSPDNLNKQVKNKG 303

Db 244 NVLFKRYEIIIEPRVQVKRLAEFLCEFTKEEESGVSBEILKLSLRLNLSNLEVNKG 303

QY 304 SSYNSKIDNKHFFRKGEVDMWANYLTSEMICKLE 337

Db 304 TT-RIGVDSQVFRKGEVDMWKNHLPQWAKTFD 336

RESULT 14

Q92410 PRELIMINARY; PRT; 325 AA.

AC O82410;

DT 01-NOV-1998 (TrEMBLrel. 08, Created)

DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Steroid sulfotransferase 3.

OS Brassica napus (Rapel).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC eurosids II; Brassicales; Brassicaceae; Brassica.

OX NCBI\_TaxID=3708;

QY [1] SEQUENCE FROM N.A.

RP MEDLINE=99340019; PubMed=10409637;

RX Rouleau M., Marsolais F., Richard M., Nicolle L., Voigt B., Adam G.,

RA Varin L.;

RT "Inactivation of brassinosteroid biological activity by a salicylate-

RT inducible steroid sulfotransferase from Brassica napus.";

RL J. Biol. Chem. 274:20925-20930(1999).

DR EMBL; AF000307; AAC63113.1; -

DR PIR; T07833; T07833.

DR HSSP; P50224; ICJM.

DR GO; GO:0008146; F:sulfotransferase activity; IEA.

DR GO; GO:0016740; F:transferase activity; IEA.

DR InterPro; IPR000863; Sulfotransferase.

DR Pfam; PF00685; Sulfotransfer; 1.

DR ProDom; PD001218; Sulfotransferase; 1.

KW Transferase.

SQ SEQUENCE 325 AA; 36940 MW; 2DD1BCBCEE203AE1 CRC64;

Query Match 26.5%; Score 546; DB 10; Length 325;

Best Local Similarity 40.2%; Pred. No. 8.2e-30;

Matches 123; Conservative 52; Mismatches 117; Indels 14; Gaps 7;

QY 37 DSKWYKRYREIVSSLPSPN-DYMGDTWRLYKGFVPGIMAFEDNFKARETDILIT 95

Db 13 DEKLTQTRLLISLSEKGLVSIQYQFGRWHTALQGLITCQHFKAQSDILIT 72

QY 96 LPKAGTWTWKALTFALLTRDYNHPSPTPLFPNPHSCVQNLLEYLYMGRENTMPDMLN- 154

Db 73 NPXSGTTWLKSLFVAFALINRHKFPVSSGDHPLLVTPHLLVPFMGVVY----ESPDDFS 128

QY 155 LNESPRLFAGHIPYSLLPASVLSKSGTKIINISNRKSTFVSFWKFGNLINPKD--LE 212

Db 129 LLFPRLMNTHTSHLSLPESVYKSSSQIVYCCRNPKDMFVSLMHPGKCLAPQETADYPLE 188

QY 213 KSVDPFASGISFCGPEWFOAEFTWAASNSN-LILLSYEEMLEKPVENKVKLAERFCG 271

Db 189 KAVEARFCQKFTAGFDVHLEWYASLEPNKVLFTVYEELKQKTEVEVKRIAEPICG 248

QY 272 FTDDKEKQIVDEIVKLCSPDNLNKQVKNKGSSNSKIDNKHFFRKGEVDMWANYLTSE 331

Db 249 FTAEE-----VSEIVKLCSPESLSRLVNRQKLPNG-IETNAPFRKGEIGWRDTLSES 303

QY 332 MIKKE 337

Db 304 LADAID 309

RESULT 15

Q92PQ6 PRELIMINARY; PRT; 331 AA.

AC Q92PQ6;

DT 01-MAY-1999 (TrEMBLrel. 10, Created)

DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Putative steroid sulfotransferase.

GN A2G03750

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;

QY [1] SEQUENCE FROM N.A.

RP Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,

RA Palm C.J., Theologis A., Ecker J., Davis R.W.;

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

QY [2] SEQUENCE FROM N.A.

RP Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,

RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;

RT "Full-length messenger RNA sequences greatly improve genome

RT annotation.";

RL Genome Biol. 0:0-0(2002).

QY [3] SEQUENCE FROM N.A.

RP Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,

RA Feldmann K.;

RT "Full-length cDNA from Arabidopsis thaliana.";

RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF325065; AAK17133.1; -

DR EMBL; AY087921; AAM65471.1; -

DR PIR; H84451; H84451.

DR HSSP; P49891; 1AQU.

DR GO; GO:0008146; F:sulfotransferase activity; IEA.

DR GO; GO:0016740; F:transferase activity; IEA.

DR InterPro; IPR000863; Sulfotransferase.

DR Pfam; PF00685; Sulfotransfer; 1.

DR ProDom; PD001218; Sulfotransferase; 1.

KW Transferase.

SQ SEQUENCE 331 AA; 38400 MW; 22E0E0A2D8CF1154 CRC64;

Query Match 26.5%; Score 545; DB 10; Length 331;

Best Local Similarity 38.4%; Pred. No. 9.9e-30;

Matches 118; Conservative 61; Mismatches 118; Indels 10; Gaps 8;

QY 36 EDSKRYKRYREIVSSLPSPN-DYMGDTWRLYKGFVPGIMAFEDNFKARETDILIT 94

Db 15 KDNVSQETKNLITSLPSDKDFMGYGLNYKGCWYYPNTLQAVLDVQKHPRDITDILA 74

QY 95 TLPKAGTWTWKALTFALLTRDYNHPSPTPLFPNPHSCVQNLLEY-LYMGRENTMPDMLN- 153

Db 75 SLPGGTTWLKSLFPAVVRKRGRTQTPTPLLPNPHDLVPFLEVELYANSQ--IPDLA 132

QY 154 MLNESPRLFAGHIPYSLLPASVLSKSGTKIINISNRKSTFVSFWKFGNLINPKD--LDL 211

Db	133	KYS-SPMIFSTHMLQALREATTKA-CKTVVCRGIGKIDFVSGWHYRNMHLRTKMDQATF	190
Qy	212	EKSVDFIFASGISFCGPENNFQAEF-TNAASTNSNLLLSYEEMLEKPVENVKLAEFMGC	270
Db	191	ELMFDAYCRGVLLYGPYWEHVLGYWKGSLKAKENLFLMKYEEIEEPVQVKRLAEFLC	250
Qy	271	GFTDDREKQGI VDEIVKLCSPDNLKQOQVKNKGSSYNSKIDNKHFFRKGEVRDWMNYLTS	330
Db	251	PFTKEEESGSVEEILKCLSLRNLNSLNVKNKGT-RIGVDSQVFFRKGEVGDWKNHLP	309
Qy	331	EMIKKLE	337
Db	310	QVAKTFD	316

Search completed: March 20, 2004, 01:38:31  
Job time : 38 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 20, 2004, 03:44:29 ; Search time 3111 Seconds  
(without alignments)  
3526.461 Million cell updates/sec

Title: US-09-854-122-16  
Perfect score: 2059  
Sequence: 1 TRGITGIAVACLPLIMAGIL.....CLLSYSFVNNFILLKKKK 391

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Zgapop 6.0 , Zgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xih  
-Q=/cgn2\_1/USPTO.spool/US09854122/runat\_18032004\_100000\_3261/app\_query.fasta\_1.583  
-DB=EST -QPM=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09854122@cgn\_1\_2810 @runat\_18032004\_100000\_3261 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFBLOCKS=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_estl:\*  
10: gb\_estc2:\*  
11: gb\_htc:\*  
12: gb\_est4:\*  
13: gb\_est3:\*  
14: gb\_est5:\*  
15: gb\_estfun:\*  
16: em\_eston:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_man:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rod:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gse1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	522.5	25.4	843	29	CC701537	CC701537 OG0BJ24TV
C	510	24.8	831	28	EZ797778	EZ797778 PUFZ85TB
3	501	24.3	1137	14	CK209284	CK209284 FGAS02104
4	499.5	24.3	975	29	CG290240	CG290240 OGZAR72TV
C	485.5	23.6	990	29	CC698091	CC698091 OG0BP20TV
C	482	23.4	859	14	CF833911	CF833911 UCRCS02_0
C	471	22.9	870	29	CC691394	CC691394 OGUEY95TV
C	471	22.9	1010	29	CC694910	CC694910 OGUBW78TV
C	469.5	22.8	934	29	CG919663	CG919663 MBET65TR
10	469	22.8	950	10	BF065941	BF065941 HV_CED001
11	466.5	22.7	1118	14	CB330744	CB330744 SpEST709
C	451	21.9	877	29	CC614744	CC614744 OGUGO80TV
13	436	21.2	807	29	CG196721	CG196721 PUFZ85TB
14	432	21.0	855	29	CG958535	CG958535 MBEGO3TF
15	426.5	20.7	966	29	CG289067	CG289067 OGKCR72TV
C	425	20.6	872	29	CG848241	CG848241 ZMMBB5031
17	412	20.0	819	28	BZ490661	BZ490661 BONIS44TR
18	411	20.0	769	14	CF554514	CF554514 GGN98 Inf
19	411	20.0	843	29	CG300068	CG300068 OGXF123TV
20	410	19.9	694	13	BUS34467	BUS34467 T061E09_P
21	407.5	19.8	738	14	CF418052	CF418052 UODA-PP_1
C	407	19.8	757	28	EZ062934	EZ062934 LLC0281L
22	405.5	19.7	698	13	BUS31310	BUS31310 T020A09_P
C	403.5	19.6	717	28	BZ080154	BZ080154 11a58c11
24	403.5	19.6	725	28	BH549521	BH549521 BORCD19TF
25	403.5	19.6	731	14	CA990082	CA990082 EST643590
26	402	19.5	982	29	CC602680	CC602680 OGVEY34TC
27	402	19.5	795	14	CA799562	CA799562 sat35810
28	400.5	19.5	580	10	BE942160	BE942160 EST421739
29	398.5	19.4	811	28	BH733391	BH733391 BOMIU60TR
30	396	19.2	869	29	CG848310	CG848310 ZMMBB5031
31	395	19.2	1074	14	CK213074	CK213074 FGAS02497
C	393	19.1	630	14	CA990182	CA990182 EST643690
33	392	19.0	828	28	BH705401	BH705401 BOMBB06TF
34	391.5	19.0	895	29	CG352888	CG352888 OGOF115TH
C	391	18.9	734	28	CC347438	CC347438 OGMBM41TV
36	388.5	18.9	781	14	CA782758	CA782758 sat54511
37	387.5	18.8	624	14	CD427868	CD427868 ETH1_34_G
38	386.5	18.8	624	14	CD427868	CD427868 ETH1_34_G
39	385	18.7	689	12	BG600378	BG600378 EST505273
40	383.5	18.6	656	14	CF810624	CF810624 CA327F_CD
41	383	18.6	1164	14	CK171170	CK171170 FGAS04631
42	382	18.6	1098	14	CK207849	CK207849 FGAS01951
43	381.5	18.5	610	14	CF810628	CF810628 CA330F_CD
44	380	18.5	786	10	BF275081	BF275081 GA_EB002
45	379.5	18.4	699	14	CD862457	CD862457 AZ01_103J

ALIGNMENTS

RESULT 1  
CC701537  
LOCUS  
DEFINITION OG0BJ24TV ZM.0.7.1.5 KB Zea mays genomic clone ZMMBMA034D24,  
genomic survey sequence.  
ACCESSION CC701537  
VERSION CC701537.1 GI:32106313  
KEYWORDS GSS.  
SOURCE Zea mays  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

```

REFERENCE 1 (bases 1 to 843)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
         Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
         Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
TITLE Consortium for Maize Genomics
JOURNAL Unpublished (2002)
COMMENT Other_GSSs: OGOBJ24TH
         Contact: Cathy Whitelaw
         TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA
         Tel: 301-838-5843
         Fax: 301-838-0208
         Email: whitelaw@tigr.org
         Seq primer: TR
         Class: sheared ends.
FEATURES
    source
        location/Qualifiers
            1..843
            /organism="Zea mays"
            /mol_type="genomic DNA"
            /strain="B73"
            /db_xref="taxon:4577"
            /clone_lib="ZMBMa0344D24"
            /clone_lib="ZM_0.7_1.5_KB"
            /note="Vector: pBGSK; Site 1: HindII; 0.7-1.5 kb
            methylation filtered genomic DNA library"
ORIGIN
Alignment Scores:
Pred. No.: 4,016-47 Length: 843
Score: 522.50 Matches: 111
Percent Similarity: 57.53% Conservative: 57
Best Local Similarity: 38.01% Mismatches: 99
Query Match: 25.38% Indels: 25
DB: 29 Gaps: 7
US-09-854-122-16 (1-391) x CC701537 (1-843)
QY 56 TyrTrpGlyAspThrMetArgLeuTyrLysGlyPheTrpGlnMetGlyTyrLeuValPro 75
   ::::::::::::::::::::
DB 2 TTTTGGCTCGGTGAGTACGATTAAGGAAGGC-----GTCCCC 40
QY 76 GlyIleMetAlaPheGluAspAsnPhelYsAlaArgGluThrAspIleLeuThrThr 95
   ::::::::::::::::::::
DB 41 GCGGT-----CATTCGTCTCGAGCCGAGACCCACCGAGTGTCTCCCAAGC 91
QY 96 LeuProLysAlaGlyThrTrpThrLysAlaLeuThrPheAlaIleLeuThrArgAsp 115
   ::::::::::::::::::::
DB 92 TTTCCCAAGTCTGGCACCACCTGGCTCAAGGCCCTCGCTTCGCGACGCTGAAGCGTTCC 151
QY 116 ValAsnHisProSerSerProThrHisProLeuLeuPheAsnProHisSerCysVal 135
   ::::::::::::::::::::
DB 152 AGCATCCGCGCTGTGACGACGACCATCCGCTCGCGCATTCGCAACCTCATGACTGTCTC 211
QY 136 GlnAsnLeuGluTyrLeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeu 155
   ::::::::::::::::::::
DB 212 AGGTCTCTGAACTTCAATTCACACGACGAGAGGATGACTCGAGCGCTC----- 262
QY 156 AsnGluSerProArgLeuPheAlaGlyHisIleProTyrSerLeuLeuProAlaSerVal 175
   ::::::::::::::::::::
DB 263 ---CCGTCCCGCGTGTGTGGCGACGACCTCCCTCTCCCTGTGTGCGCGCAGCATC 319
QY 176 Leu-----LysSerGlyThrLysIleIleAsnIleSerArgAsnArgLysSer 191
   ::::::::::::::::::::
DB 320 ACCGGGAGGAGGAGCGCTCGGGGTCCGGATCGTCTACGTTTGGCGGAACCCAGGAC 379
QY 192 ThrPheValSerPheTrpLysPheGlyAsnLeuIleAsnProAspLysLeuAsp--- 210
   ::::::::::::::::::::
DB 380 GTGCTGTCTCCCTCGGTGTTCACGAGAAAGGACCGCGCTGTGGGGTTGACGCC 439
QY 211 -----LeuGluLysSerValAspIlePheAlaSerGlyIleSerPheCysGly 226
   ::::::::::::::::::::
DB 440 CGGTGCTTACGATCAGGAGGCGCTTGGAGCTGTCTTGTGTGCGCGCTGTCTCGCGGC 499

```

227 ProGluTrpAsnPheGlnAlaGluPhe---ThrAsnAlaAlaSerThrAsnSerAsnLeu 245  
 500 CCACACTGGAACACACCGCTCTCCACTACTGGAAGAGAGCGTGAGGAGGCTGCACAGGGTG 559  
 246 LeuLeuLeuSerTyrGluGluMetLeuGluLysProValGluAsnValLysLysLeuAla 265  
 560 TTGTTCTTCTGTTACAGAGGAGTGTGATTGAGCCGAGCGGCGGCTGAGAAAGCTGGCC 619  
 266 GluPheMetGlyCysGlyPheThrAspGluGluLysGlnGlyLeuValAspGluIle 285  
 620 AATTTCATGGGCTGTGATTCTCAGAGGAGGAGGAGCGGGGTGTGTGAGCGGCATC 679  
 286 ValLysLeuCysSerPheAspAsnLeuLysAsnGlnValAsnLysAsnGlySerSer 305  
 680 GTGAGCTGTGTAGCTTGGGAGCTCAGAGACATGAGAGTAAATAGAATGAGAGCACT 739  
 306 TyrAsnSerLysIleAspAsnLysHisPheArgLysGlyGluValArgAspTrpAla 325  
 740 TTGTTGGGG---ATCAAGAACGAAAGCTTCTTCAGGAGGAGAGTGTCTGGGAGCTGGAGC 796  
 326 AsnTyrLeuThrSerGluMetIleLysLysLeuGlu 337  
 797 AAGCATATGACGAGGAGATGGCGCAGAGGCTAGAC 832

RESULT 2  
 BZ797778/c 831 bp DNA linear GSS 17-MAR-2003  
 LOCUS PUF285TB ZM 0.6\_1.0\_KB Zea mays genomic clone ZMBTa317002,  
 DEFINITION genomic survey sequence.  
 ACCSSION BZ797778  
 VERSION BZ797778.1 GI:29000480  
 KEYWORDS GSS.  
 SOURCE Zea mays  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.  
 1 (bases 1 to 831)  
 Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,  
 Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and  
 Bennetzen,J.  
 Maize Genomics Consortium  
 Unpublished (2003)  
 Other\_GSSs: PUF285TD  
 Contact: Cathy Whitelaw  
 TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-5843  
 Fax: 301-838-0208  
 Email: whitelaw@tigr.org  
 Seq primer: TR  
 Class: sheared ends.  
 Location/Qualifiers  
 1..831  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone\_lib="ZMBTa317002"  
 /clone\_lib="ZM\_0.6\_1.0\_KB"  
 /notes="Vector: pCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high  
 Cot selected genomic DNA library"

Alignment Scores:  
 Pred. No.: 9,596-46 Length: 831  
 Score: 510.00 Matches: 108  
 Percent Similarity: 57.14% Conservative: 52  
 Best Local Similarity: 38.57% Mismatches: 104  
 Query Match: 24.77% Indels: 16  
 DB: 28 Gaps: 6  
 US-09-854-122-16 (1-391) x BZ797778 (1-831)





```

Db      887 AGAGACATGGAGGTAATAGAAATGAGACACTTTGTTGGG---ATCAAGAACGAAAGC 943
QY      315 Phe-PheArgLysGlyGluValAlaArgAspTTP 324
Db      944 TTCTTTCAGGAAGGAGTGTGTTGGGACTGG 974

RESULT 5
LOCUS   CC698091/c
DEFINITION OGBP20TV_ZM_0.7_1.5_KB_Zea_mays_genomic_clone_ZMMBMA0347C16,
genomic survey sequence.
ACCESSION CC698091
VERSION   CC698091.1 GI:32102867
KEYWORDS  GSS.
SOURCE    Zea mays
ORGANISM  Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 990)
Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,
Citek, R.W., Nurnberg, A., Robbins, D. and Lakey, N.
Consortium for Maize Genomics
Unpublished (2002)
Other GSSs: OGBP20TH
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TP
Class: sheared ends.

FEATURES
    source
        location/Qualifiers
            1..990
                /organism="Zea mays"
                /mol_type="genomic DNA"
                /strain="B73"
                /db_xref="taxon:4577"
                /clone_ZMMBMA0347C16"
                /clone_lib="ZM 0.7 1.5 KB"
                /note="Vector: pBCKS-; Site 1: HincII; 0.7-1.5 kb
                methylation filtered genomic DNA library"

ORIGIN
Alignment Scores:
Pred. No.:      6.29e-43      Length:      990
Score:          485.50      Matches:    121
Percent Similarity: 51.59%      Conservative: 41
Best Local Similarity: 38.54%      Mismatches: 127
Query Match:     23.55%      Indels:     25
DB:              29          Gaps:         9

US-09-854-122-16 (1-391) x CC698091 (1-990)

QY      61 MetArgLeuTyrLysGlyPheTyrPgluValProGlyLeuMetAlaPhe 80
Db      998 CTGCGCCCTTACAGAGGCTTCTGGTCCGGAGCAGCTGGGTGCCGGGACCATCGTGTTC 929
QY      81 GluAspAsnPheLysAlaArgGluThrAspIleLeuThrLeuProLysAlaGly 100
Db      928 CAGCGCCGCTTACACCGCGCCCGCCGACGACGTGATCTCTCGGAGCTACCCCAAGTGGCG 869
QY      101 ThrThrTTPThrLysAlaLeuThrPheAlaIle-LeuThrArgAspValAsnHisProSe 120
Db      858 ACCAGTGGCTGAAGCGCTGGCTTCGCCACGGCGCGGTACGGGTACCGCGCGG 809
QY      120 rSerProThrHisProLeuLeuPheAsnProHisSerCysValGlnAsnLeuGluTy 140
Db      808 TGGCGCCAGGACCGCGCTCCGACGGCTCAACCCGACGACTCGCTCCGTTTCATCGACGA 749
QY      140 rleuTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsnGluSerProAr 160

```

```

Db      748 GATCTTCGCCAGCGGGAGGAGCG-----AAGCTCGAGGCCCTC---CGCTGCCCGAG 698
QY      160 glenPheAlaGlyHisIleProTyrSerLeuLeuProAlaSerVal-----LeuLysSe 178
Db      697 GCTGATGAACACGACCTGCGCTAGCGCTGCTCCCGCGCCCGCTACCGCCACCGCCAC 638
QY      178 rGlyThrLysIleLeuAsnIleSerArgAsnArgLysSerThrPheValSerPheTyrIly 198
Db      637 CGGTGCAGGTTCGGTACGTGTGCAGGAGACCCCAAGGACATGGTGGTCTGCTCTGGCA 578
QY      198 sPheGlyAsnLeuIleAsnProAspLysLeuAspLeuGluLysSerValAspIlePh 218
Db      577 CTTCCTCCGCGCGCAAGCGACCTCTTGTTC-----GCCGACACGT 533
QY      218 eAlaSer-----GlyIleSerPheCysGlyProGluTTPAsnPheGlnAlaGluPh 235
Db      532 CGAGTCCGTCCTCGACGGACCGTGTGTGGCCCGGTCTGGGACCACGCTCTCTCTCTA 473
QY      235 eThrAsnAlaAla----SerThrAsnSerAsnLeuLeuLeuSerTyrGluGluMetLe 254
Db      472 CTGGCGTGGAGCGTGGCGGCCCGGACAGGCGCGCTCTTCTCAGGTACGAGGACATGCT 413
QY      254 uGluLysProValGluAsnValLysLeuAlaGluPheMetGlyCysGlyPheThrAs 274
Db      412 GCGGACCCCGCGCGACGTCGCGGGCTGCCAGTTCATGGCGCGCCGCTTCTCGGC 353
QY      274 pAspGluGluLysGlnGlyIleValaspGluIleValLysLeuLysCysSerPheAspAsnLe 294
Db      352 CGCGAGGAGGCGCGCGAGCTCGCGCCCTCGTGAGCTCTCGAGCTTCGACGTTCCAGCAGAT 293
QY      294 uLysAsnGlnGlnValAsnLysAsnGlySerSerTyrAsnSerLys-----IleAs 311
Db      292 GAAGGCTGGAGGTGAACAGCCCTGTAGCGGACCGCGGGAGGTACCGGCCCATGCG 233
QY      311 pAsnLysHisPhePheArgLysGlyGluValArgAspTTPAlaAsnTyrLeuThrSerGl 331
Db      232 GCGCGACGCGCTTCTTCGGAAGGCTGTCGCCGGGACTGGCGGAAACCAATGACGCGGA 173
QY      331 uMetIleLysLysLeuGluThrAlaGlyLysIleAsnGlu-SerGluLysHisLeuLeuS 351
Db      172 GATGGCGCGCGCTGGAC-----GGGATCTTCGTTGAAAGTCCCAAG----- 129
QY      351 exAsnLysAsnLeuThrAsnPheAsnLeuAsnAsnTyr 363
Db      128 --GGACGGGGCTTATCTTCTCATATGCAATGTGTAT 93

RESULT 6
CF833911/c
LOCUS   CF833911
DEFINITION UCRS02_04105 f Ruby Orange Ovary at Anthesis cDNA Library Citrus
sinensis cDNA clone CS_REA04105, mRNA sequence.
ACCESSION CF833911
VERSION   CF833911.1 GI:38036870
KEYWORDS  EST.
SOURCE    Citrus sinensis
ORGANISM  Citrus sinensis
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosids; eurosida II; Sapindales; Rutaceae; Citrus.
1 (bases 1 to 859)
Close, I.J., Roose, M.L., Federici, C.F., Fenton, R.D., Wanamaker, S.,
Kim, H.R., Kudrna, D., Wing, R. and Yu, Y.
Development of EST Resources and New Genetic Markers for California
Citrus - Ruby Orange Ovary at Anthesis
Unpublished (2003)
Contact: Timothy Close
Department of Botany & Plant Sciences, University of California
Riverside, CA, 92521-0124
Tel: 9097873318
Fax: 9097874437
Email: timothy.close@ucr.edu
Seq primer: T7.

```





/clone\_lib="ZM 0.7.1.5\_KB"  
/note="vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb  
methylation filtered genomic DNA library"

## ORIGIN

## Alignment Scores:

Pred. No.: 2,16e-41 Length: 870  
Score: 471.00 Matches: 101  
Percent Similarity: 55.72% Conservativeness: 50  
Best Local Similarity: 37.27% Mismatches: 104  
Query Match: 22.88% Indels: 16  
DB: 29 Gaps: 6

US-09-854-122-16 (1-391) x CC691394 (1-870)

```
QY 47 lIeValSerSerLeuProSerAsnAspTyrTrpGlyAsp---ThrMetArgLeuTyrLys 65
D 815 ATCATTCCTTCGCTGCGCCCTCGAGACCAAGGTCGACCGTTCCTCCGACCGGTAGCC 756
QY 66 GlyPheTrpGlnMetGlyTyr---LeuValProGlyIleMetAlaPheGluAspAsnPhe 84
D 755 AATTTTGGCTGCTGAGTAGCATTGAAGAGAGGCTCCCGCGTTCATTCTCGCTTC 696
QY 85 LysAlaArgGluThrAspIleIleLeuThrLeuProLysAlaGlyThrTrpThr 104
D 695 GAGCGGAGACCCAGCGAGTCTTCGCAAGCTTCCCAAGTCTCCGACCACTGCTGCTC 636
QY 105 LysAlaLeuThrPheAlaIleLeuThrArgAspValAsnHisProSerProThrHis 124
D 635 AAGCCCTTCGCTTCGCGACCTGAAGGTTCCAGCATCGCGCTTCGAGGAGCAT 576
QY 125 ProLeuLeuPhePheAsnProHisSerCysValGlnAsnLeuGluTyrLeuTyrMetGly 144
D 575 CCGCTGCGCCATTGCAACCTCATGCTGCTCAGGTTCTCGAAGTCTGATTTCAACAG 516
QY 145 ArgGluThrMetProAspLeuAspMetLeuAsnGluSerProArgLeuPheAlaGly 164
D 515 CAGAGGATGAGCTCGAGGCGCTC-----CGTCCCGCGTGTGCTGGCCACG 468
QY 165 HisIleProTyrSerLeuLeuProAlaSerValLeu-----LysSerGlyThr 180
D 467 CACTCCCTACTCCTGCTGCGGAGCATCACCGGCGAGCAGGAGCGCTCGGGTGC 408
QY 181 LysIleIleAsnIleSerArgAsnArgLysSerThrPheValSerPheTrpLysPheGly 200
D 407 CGGATGCTTACGTTTGCCTGCGGAAACCCCAAGGACGTGCTGCTCTCTGCTGCTTCCAG 348
QY 201 AsnLeuIleAsnProAspLysLeuLeuAsp-----LeuGluLysSerVal 215
D 347 AGAAGGAGCGCGCGCTGTGGGTTGACCCCGGTGTTACGATCCAGAGGCGCTG 288
QY 216 AspIlePheAlaSerGlyIleSerPheCysGlyProGluTrpAsnPheGlnAlaGluPhe 235
D 287 GAGCTGTCTGTGATGGCGCTGTCTCGCGCGCCACAGTGGAAACCACTCTCCACTAC 228
QY 236 ---ThrAsnAlaAlaSerThrAsnSerAsnLeuLeuLeuSerTyrGluGluMetLeu 254
D 227 TGGAGGAGAGCGTCAGAGGCGCTGACAGGCTGTGTTCTCTCTACGAGGAGTCTG 168
QY 255 GluLysProValGluAsnValLysLysLeuAlaGluPheMetGlyCysGlyPheThrAsp 274
D 167 ATTGAGCCGAGCGCGACGTCAGAAAGCTGCGCAATTTTCATGGGCTGTGGATTCTCAGAG 108
QY 275 AspGluGluLysGlnGlyIleValAspGluIleValLysLeuCysSerPheAspAsnLeu 294
D 107 GAGGAGGAGGAGCGCGGGTGTGAGCCCATCTGTTGAGCTGTGTAGCTTGGGAAATG 48
QY 295 LysAsnGlnGlnValAsnLysAsnGlySerSer 305
D 47 AGAGACATGGAGGTAATAAGAAATGGAAGCACT 15
```

RESULT 8

CC629410/c

LOCUS CC629410 1010 bp DNA linear GSS 19-JUN-2003  
DEFINITION OGUBW78TV ZM 0.7.1.5\_KB Zea mays genomic clone ZMMBMA0404M11,  
genomic survey sequence.  
ACCESSION CC629410  
VERSION CC629410.1 GI:32001744  
KEYWORDS GSS.  
SOURCE Zea mays  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE 1 (bases 1 to 1010)  
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,  
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,  
Citek,B.W., Nunberg,A., Robbins,D. and Lakey,N.  
CONSORTIUM for Maize Genomics  
Unpublished (2002)  
Other\_GSSs: OGUBW78TH  
Contact: Cathy Whitelaw

## TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TF

Class: sheared ends.

## FEATURES

## source

1..1010  
Location/Qualifiers  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBMA0404M11"  
/clone\_lib="ZM 0.7.1.5\_KB"  
/note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb  
methylation filtered genomic DNA library"

## ORIGIN

## Alignment Scores:

Pred. No.: 2,62e-41 Length: 1010  
Score: 471.00 Matches: 115  
Percent Similarity: 55.74% Conservativeness: 55  
Best Local Similarity: 37.70% Mismatches: 122  
Query Match: 22.88% Indels: 14  
DB: 29 Gaps: 8

US-09-854-122-16 (1-391) x CC629410 (1-1010)

```
QY 42 LysArgTyrArgGluIleValSerSerLeuPro-SerAsnAspTyrTrpGlyAspThrMe 61
D 988 AAGAGCTACGGCGACCTGGTATCCACTCTCCGGGCTAGAGCGGTGGGTATCGTCTCGT 929
QY 61 tArgLeuTyrLysGlyPheTrp---GlnMetGlyTyrLeuValProGlyIleMetAlaPh 80
D 928 GCAG---TACCAGAACTACTGCTGAACCCCGGAGGCTCCA-GAAACATATCCCCCGT 873
QY 80 eGluAspAsnPheLysAlaArgGluThrAspIleIleLeuThrThrLeuProLysAlaGl 100
D 872 GAAAGAGCTCTACAGCCTCGCGCCGACGACATCTCTCGCACCTACTCTGAAATCGCG 813
QY 100 YThrThrTrpThrLysAlaLeuThrPheAlaIleLeuThrArgAspValAsnHisPro-- 119
D 812 GACCACTGGCTGAAGCGCTGCGCTTCGCCATCAGACCCGCGCGCCGCCACCCAGGC 753
QY 120 ---SerSerProThrHisProLeuLeuPhePheAsnProHisSerCysValGlnAsnLe 138
D 752 CTTGCGCGCGCGGACACCCCTGCTCACCCTGTCACCCGCGAGGAGCGTCCCGCACT 693
QY 138 uGluTyrLeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsnGluSe 158
D 692 GGAGGTGCACACGCCCGAGGCGAG-----GGCCTGGCAGACATCGAGAGGCTG---CCGTC 642
QY 158 rProArgLeuPheAlaGlyHisIleProTyrSerLeuLeuProAlaSerValLeuLysSe 178
```

```
Db 641 GCCAGGCTGCTCGGAGCAGCAGCTGCGCCCTCTCCCTGCTCCGCGCGCGCTGCGCACCTC 582
Qy 178 rGlyThrLysIleIleAsnIleSerArgAsnArgLysSerThrPheValSerPheTrpLys 198
Db 581 CGGCTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 522
Qy 198 sPheGlyAsnLeuIleAsnProAspLysLeu---LeuAspLeuGluLysSerValAspL 217
Db 521 CTTGCGTGAAGGCGATCGCGGGGCGAGGTCCCGCGTGGAGCTGGACGCGCGCTGCCCAT 462
Qy 217 ePheAlaSerGlyIleSerPheCysGlyProGluTrpAsnPheGlnAlaGluPhe---Th 236
Db 461 GTTCTCGAGGGGCTGCGCTGCTGCGGCGCGCTGCGGAGCTACCTCGAGTACTGGAA 402
Qy 236 rAsnAlaAlaSerThrAsnSerAsnLeuLeuLeuSerTyrGluGluMetLeuGluLys 256
Db 401 GGAGAGCTGGCGAGCGCTGAGCGGTCTCTTTTCTCAGGTACGAGGAGATGCTGCGGA 342
Qy 256 sProValGluAsnValLysLysLeuAlaGluPheMetGlyCysGlyPheThrAspAspL 276
Db 341 CCGGCTGCGTGGCTGAGGAGCTGCGCGCTCTTTCGCGCTGCGCTGCGCTGCGGAGGA 282
Qy 276 uGluLysGlnGlyIleValAspGluIleValLysLeuLysSerPheAspAsnLeuLysAs 296
Db 281 AGAGGGCGAGGGTCCCGAGGAGATGTGAGGCTGTGAGCTGCGAGATGCTGAGCGG 222
Qy 296 nGlnValAsnLysAsnGlySer-----SerTyrAsnSerLysIleAspAsnLys 313
Db 221 CTTGGAAGAGCATCGGACCGGAGATCTTGATTCGGGAGACACACCGTGTGCGGAATC 162
Qy 313 sHsPheArgLysGlyGluValArgAspTrpAlaAsnTyrLeuSerGluMetI 333
Db 161 TACTTCTTCAGAGGGCGAGGTTCGGGACTGGGAGACCATCTTGACAAAGAGATGGG 102
Qy 333 eLysLysLeuGlu 337
Db 101 CAAGAGGCTCGAC 89
```

## RESULT 9

CG919663/c  
LOCUS MBEBT65TRB mth2 Medicago truncatula genomic clone 22L10, genomic survey sequence.  
CG919663  
CG919663.1 GI:39779346

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

Medicago truncatula (barrel medic)  
Medicago truncatula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

1 (bases 1 to 934)  
Town, C.D., Shetty, J., Koo, H. and Feldblyum, T.F.  
Sequencing of BAC ends from Medicago truncatula  
Unpublished (2003)  
Other GSSEs: MBEBT65TF  
Contact: Chris Town  
TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org

Seq primer: CAGGAACAGCTATGACC  
Class: BAC ends.

## FEATURES

## source

1..934  
/organism="Medicago truncatula"  
/mol\_type="genomic DNA"  
/cultivar="genotype Al7"  
/db\_xref="taxon:3880"  
/clone="22L10"

RESULT 10  
BF065941

## ORIGIN

## Alignment Scores:

Pred. No.: 3,486-41 Length: 934  
Score: 469.50 Matches: 99  
Percent Similarity: 59.70% Conservative: 58  
Best Local Similarity: 37.64% Mismatches: 99  
Query Match: 22.80% Indels: 7  
Gaps: 5

US-09-854-122-16 (1-391) x CG919663 (1-934)

```
Qy 34 LysGluGluAspSerLysMetTyrLysArgTyrArgGluLeuValSerSerLeuProSer 53
Db 780 AAAGAGATGAC-----TTAGGCAAGAAATGCAAGAGCTTGATACAAACCTTGCCATTA 727
Qy 54 AsnAspTyrTrpGlyAsp---ThrMetArgLeuTyrLysGlyPheTrpGlnMetGlyTyr 72
Db 726 GTAGAGGGTGGATCGATCCACCTTTCATGAATATCAAGAGATTTTGGTTGCTCCCAAG 667
Qy 73 LeuValProGlyIleMetAlaPheGluAspAsnPhelLysAlaArgGluThrAspIleLeu 92
Db 666 ATTCTCCAAAGTGCTTTATCTTTTCAAAAGCAGCTTTCAGGCCATTTGATACAGATATTATC 607
Qy 93 LeuThrThrLeuProLysAlaGlyThrTrpThrLysAlaLeuThrPheAlaIleLeu 112
Db 606 TTAGTTCAATCCCAATCAGGTACCACTTGGCTCAAGCATTTGCTTTGCTTGATA 547
Qy 113 ThrArgAspValAsnHisProSerSerProThrHisProLeuLeuPhePheAsnProHis 132
Db 546 AACCGTAACAAATATCCAAATATTCATAAAATCATCTTTTACTCAGCTTCCAAACCCCTCAT 487
Qy 133 SerCysValGlnAsnLeuGluTyrLeuTyrMetGlyArgGluAsnThrMetProAspLeu 152
Db 486 GTTCTTGTTCCTTCATGAGATTAACCTTTACTATGAACTGATATTTCTTGAAGCTC 427
Qy 153 AspMetLeuAsnGluSerProArgLeuPheAlaGlyHisIleProTyrSerLeuLeuPro 172
Db 456 AATTCACATATCC---CCTCCAAAGAAATCTTCCACACACAAATCCTTATGTTTGTGCCA 370
Qy 173 AlaSerValLeuLysSerGlyThrLysIleLeuAsnIleSerArgAsnArgLysSerThr 192
Db 369 AAATCTTTGAAAGAGTCCAACTGTAAAGTGTGTATCTTTGTAGAGATCTCTTAAGACACT 310
Qy 193 PheValSerPheTrpLysPheGlyAsnLeuIleAsnProAspLys-----LeuLeuAsp 210
Db 309 TTTGTTTTCATTTTGGCATTTCTCAACCAAGCTAAAGAGAACAAAGTAGAGGACACTCCA 250
Qy 211 LeuGluLysSerValAspIlePheAlaSerGlyIleSerPheCysGlyProGluTrpAsn 230
Db 249 TTAGAAGAGGCTTTTGAGAGTTTGTAGAGGAAACAACTTCTTATGGACCCCTTTTGGAT 190
Qy 231 PheGlnAlaGluPhe---ThrAsnAlaAlaSerThrAsnSerAsnLeuLeuLeuSer 249
Db 189 CATGTATTAGGATACCTGGAGGAAAGCTTGGAGAGACCGGAGAGAGGTATGTTTGAAGA 130
Qy 250 TyrGluGluMetLeuGluLysProValGluAsnValLysLysLeuAlaGluPheMetGly 269
Db 129 TATGAAGAAATGAAATGAAACCTTAATTTGTTTGAAGAGAGATTCGTAAGTTTGTGGG 70
Qy 270 CysGlyPheThrAspAspGluLysGlnGlyIleValAspGluIleValLysLeuCys 289
Db 69 TGTCCATCTCCGAGAGAGAGAAATCCCAATGCTGTGTGTGATGATATATTACATTGTGT 10
Qy 290 SerPheAsp 292
Db 9 AGTTTGTAG 1
```

LOCUS  
 DEFINITION HV\_CEB0014G16f Hordeum vulgare seedling green leaf EST library  
 HVCDNA0005 (Blumeria challenge) Hordeum vulgare subsp. vulgare  
 CDNA clone HV\_CEB0014G16f, mRNA sequence.

ACCESSION  
 VERSION BF065941  
 KEYWORDS EST.  
 SOURCE BF065941.1 GI:10842580  
 ORGANISM Hordeum vulgare subsp. vulgare  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Pooidae; Triticeae; Hordeum.  
 1 (Bases 1 to 950)  
 WING, R., CLOSE, T.J., Kleinhofs, A., Wise, R., Wei, F., Begum, D.,  
 Frisch, D., Yu, Y., Henry, D., Palmer, M., Rambo, T., Simmons, J.,  
 Choi, D.W., Fenton, R.D., Oates, R. and Main, D.  
 Development of a genetically and physically anchored EST resource  
 for barley genomics: Blumeria infected incompatible (Mla6) seedling  
 leaf cDNA library  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Wing RA  
 Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 7288  
 Fax: 864 656 4293  
 Email: rwing@clemson.edu  
 Total hg bases = 574  
 Seq primer: AATTAAACCTCACTAAAGGG  
 High quality sequence stop: 644.  
 Location/Qualifiers  
 1..950  
 /organism="Hordeum vulgare subsp. vulgare"  
 /mol\_type="mRNA"  
 /cultivar="C116151 (Mla6)"  
 /sub\_species="vulgare"  
 /db\_xref="taxon:112509"  
 /clone="HV\_CEB0014G16f"  
 /tissue\_type="seedling green leaf"  
 /lab\_host="SOLR"  
 /clone\_lib="Hordeum vulgare seedling green leaf EST  
 library HVCDNA0005 (Blumeria challenge)"  
 /note="vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;  
 C.I. 16151 (Mla6) plants were greenhouse grown in the R  
 Wise lab at Iowa State University, Ames, IA; 7 day old  
 green seedlings were challenged with isolate 5974  
 (AvrMla6) of Blumeria graminis f. sp. hordei, and leaves  
 were harvested 20 and 24 hr post-inoculation and snap  
 frozen; unanoculated leaves were harvested 20 hr  
 post-inoculation (Wei, Wise). In the TJ Close lab at the  
 University of California, Riverside, total RNA was  
 prepared from each sample pool, equal quantities of all  
 three RNA pools were combined, poly(A) RNA was purified  
 from the mixture, one primary unamplified cDNA library was  
 made, and 1 million pfu were in vivo excised to give  
 pBluescript SK(-) cDNA phagemids (Choi, Close). Phagemids  
 were plated and picked at the Clemson University Genomics  
 Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing).  
 plasmid DNA preparations, DNA sequencing and sequence  
 analysis were performed at CUGI (Wing, Yu, Frisch, Henry,  
 Simmons, Oates, Rambo, Main). The sequence has been  
 trimmed to remove vector sequence and contains a minimum  
 of 100 bases of phred value 20 or above. For more details  
 on library preparation and sequence analysis see  
 http://www.genome.clemson.edu/projects/barley. To order  
 this clone see http://www.genome.clemson.edu/orders Also  
 see Close TJ, Wing R, Kleinhofs A, Wise R (2001)  
 Genetically and physically anchored EST resources for  
 barley genomics. Barley Genetics Newsletter 31:29-30.  
 (http://wheat.pw.usda.gov/ggppages/bgn/31/cover.html")

## ORIGIN

## Alignment Scores:

RESULT 11

CB330744

LOCUS

DEFINITION

1118 bp

mRNA

linear

EST 07-MAR-2003

IPomoea batatas cDNA 5', 3', mRNA

SPEST709 Sweetpotato cDNA library

Pred. No.: 4,046-41 Length: 950  
 Score: 469.00 Matches: 110  
 Percent Similarity: 54.74% Conservative: 46  
 Best Local Similarity: 38.63% Mismatches: 102  
 Query Match: 22.78% Indels: 27  
 DB: 10 Gaps: 7

US-09-854-122-16 (1-391) x BF065941 (1-950)

QY 73 LeuValProGlyIleMetAlaPheGluAspAsnPheLysAlaArgGluThrAspIleIle 92  
 DB 1 CTCGGTCCG-----AATTCGGCAGCGCCGCTCTCTC 33

QY 93 LeuThrThrLeuProLysAlaGlyThrThrTrpThrLysAlaLeuThrPheAlaIleLeu 112  
 DB 34 CTCGGCAGCTTCCCAAGTCCGCGACCACTCGCTCGCTCAAGCGCTCGCTTCCCGCAGCTC 93

QY 113 ThrArgAspValAsnHisProSerProThrHisProLeuLeuPhePheAsnProHis 132  
 DB 94 CACCGCACCGACACCCGACACTCAGCTCGACCATCGCTCCGCGCGGCGTAACTCTAC 153

QY 133 SerCysValGlnAsnLeuGlu-----TyrLeuTyrMetGlyArgGluAsnThrMetPro 150  
 DB 154 GACTGGCTCGAGTCTCTCGAGGAGGCGCTTGGGGTCAATTGTCTCGCGAGCAGCGCGGC 213

QY 151 AspLeuAspMetLeuAsnGluSerProArgLeuPheAlaGlyHisIleProTyrSerLeu 170  
 DB 214 GACGTGCTCGCGGGCTGCTTTCGGCGCGCTGCTCGCCACCCACCTTACACCTAT 273

QY 171 LeuProAlaSerValLeuLys-----SerglyThrLysIleIleAsnIleSerArg 187  
 DB 274 CTGCGCAAGAGTGTACGCGCGCGCGGCGCACCGGGTGCAGATCGTGTACATCTGCCGT 333

QY 188 AsnArgLysSerThrPheValSerPheTrpLysPheGlyAsnLeu----- 202  
 DB 334 GACCCCAAGGACGCGCTCTCTCCAGTGGCTATTACCAAGAGAAATTTGGCTTGGCG 393

QY 203 -----IleAsnProAspLysLeuLeuAspLeuLysSerValAspIlePheAla 219  
 DB 394 GCGGTGTACGACACACCCGCGCCCAAGCCATACCATCGATGATGTTCGAGCTCTTCTGC 453

QY 220 SerGlyLysPheCysGlyProGluTrpAsnPheGlnAlaGluPheThrAsnAlaAla 239  
 DB 454 GACGGCGGATGCACCTCGCGCGCCCAATGGCACCATTCTCTCGCTTGGGAGCGGAGC 513

QY 240 SerThrAsn---SerAsnLeuLeuLeuLeuSerTyrGluGluMetLeuLysProVal 258  
 DB 514 AGAGGCACCCAGAGAGAGTGTCTCTCTCTCGTTACGAGAGATGCTCCGGACCGCGCG 573

QY 259 GluAsnValLysLeuAlaGluPheMetGlyCysGlyPheThrAspAspGluLys 278  
 DB 574 GGCAGCGTGTAGGAAGCTGGCGGAGTTTCATGGGCTCGCGCTTCTCCGCGGAGAGCGCG 633

QY 279 GlnGlyIleValAspLysLeuValLysLeuCysSerPheAspAsnLeuLysAsnGln 298  
 DB 634 GCGGGGCTCGCGAGACGCTCTGAGCTGTGCAGATGACACTTAAAGAACATGGAG 693

QY 299 ValAsnLysAsnGlySerSerTyrAsnSerLysIleAspAsnLysHisPhePheArgLys 318  
 DB 694 GTGAATAAGATGGACCGCAN-----AACTACGTGAAGATGAGGCTTCTTCGCGAAG 747

QY 319 -GlyGluValArgAspTrpAlaAsn---TyrLeuThrSerGluMetIleLysLeuGln 337  
 DB 748 GGGGGTGTGGGGGACTGGAGCAACCCATGACACCGGCGGCGATGGCGCGCGCTTGA 807

QY 337 :  
 DB 808 CAAGATTGGTGAG 820

```

sequence.
ACCESSION      CB330744.1 GI:28879873
VERSION        CB330744.1
KEYWORDS       EST.
SOURCE         Ipomoea batatas (sweet potato)
ORGANISM       Ipomoea batatas
               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
               asterids; lamids; Solanales; Convolvulaceae; Ipomoea.
REFERENCE      Jankowicz, J., Berenyi, M. and Burg, K.
AUTHORS        Identification of adaptation specific differences in the mRNA
TITLE          expression profile of drought tolerant and sensitive sweetpotato
JOURNAL        unpublished (2004)
COMMENT        Contact: Berenyi, M.
               Department of Biotechnology
               ARC Seibersdorf Research GmbH
               A2444 Seibersdorf, Austria
               Tel: +43 50550 3524
               Fax: +43 50550 3444
               Email: Maria.Berenyi@arcs.ac.at
PCR Primers   FORWARD: lambda Triplex 5', Screening Amplimer
               BACKWARD: lambda Triplex 3', Screening Amplimer
Insert length: 1118 Std Error: 0.00
Seq primer: lambda Triplex Sequencing Primers
POLYA=Yes.
FEATURES       Location/Qualifiers
               1..1118
               /organism="Ipomoea batatas"
               /mol_type="mRNA"
               /db_xref="taxon:4120"
               /tissue_type="Leaf"
               /lab_host="Lambda"
               /clone_lib="Sweetpotato cDNA library"
               /notes="Vector: Triplex2; Site 1: Sfi A; Site 2: Sfi B;
               Ipomoea batata (lam) L.; Smart cDNA library Construction
               kit; Clontech: PT3000-1"
ORIGIN
Alignment Scores:
Pred. No.:      9,45e-41      Length:      1118
Score:          466.50        Matches:     126
Percent Similarity: 49.22%    Conservative: 63
Best local Similarity: 32.81%  Mismatches:  151
Query Match:    22.66%       Indels:      45
DB:             14           Gaps:        10
US-09-854-122-16 (1-391) x CB330744 (1-1118)
QY      38 SerLysMetTyrLysArgTyrArgGluIleValSerSerLeuProSerAsnAspTyrTrp 57
Db      1 AGTCATCATATGCAATGATGAGACAGCAGCATATATCTCA-GAGGAAAGAAATGGTGG 59
QY      58 GlyAspThr---MetArgLeuTyrLysGlyPheTrpGlnMetGlyTyrLeuValProGly 76
Db      60 GGAGACAAACATTATGACGACATCAATGGTCTCTGGTTCTTGTCTTTTAGCCCCCAA 119
QY      77 IleMetAlaPheGluAspAsnPhelysAlaArgGluThrAspIleIleLeuThrLeu 96
Db      120 ACTGCCGAGGCTAGCCGAGTTTAAACCAGTCCCGACGACGAAATTTGAGTTGCTTT 179
QY      97 ProLysAlaGlyThrThrTrpThrLysAlaLeuThrPheAlaIleLeuThrArgAspVal 116
Db      180 CCCAAACCGGCTCCACATGGCTCAATCTCTTCTTACTCCATCATCAATCGC----- 233
QY      117 AsnHisProSerSerProThrHisProLeuLeuPhePheAsnProHisSerCysValGln 136
Db      234 -----TCTCTCTTCATCTCTTGTCAACACACACACACCCCGACCTCGTCCCT 281
QY      137 AsnLeuGlu---TyrLeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeu.155
Db      282 TTCCTAGAGATTCAGTCTATGACAGACGAGAAATCATCACTACATGTTGTTCGAGGAT 341

```

```

QY      156 AsnGluSerProArgLeuPheAlaGlyHisIleProTyrSerLeuLeuProAlaSerVal 175
Db      342 ACCACGAGTACTAGGCTTTTCGTATCCCATATCCCTTACCACCTGCTGGCAAAACCTTG 401
QY      176 LeuLysSerGlyThrLysIleIleAsnIleSerArgAsnArgLysSerThrPheValSer 195
Db      402 GAATCTCTGGGCTGCTGCTGGTTTATGTTGCCAGGAACCCAAAGACACCTTAACTTCT 461
QY      196 PheTrpLysPheGlyAsn-----LeuIleAsnProAspLysLeuLeuAspLeuGluLys 213
Db      462 CTCCTGGCATTTCCACCAATAAGTGAAGATGCCGACGAGGACATGGAGTTGGAGAA 521
QY      214 SerValAspIlePheAlaSerGlyIleSerPheCysGlyProGluTrpAsnPheGlnAla 233
Db      522 GCCGTGGAGAAATTTCTGAGTGGGACTGTTCTCTCGGACCTTACTACGACGATGTTTG 581
QY      234 GluPheThrAsnAlaAlaSerThrAsn---SerAsnLeuLeuLeuLeuSerTyrGluGlu 252
Db      582 GGTACACAGATGGCTACTTTGAATACCCCTAGTAATCTCTTTATTACTTATGAAGAG 641
QY      253 MetLeuGluLysProValGluAsnValLysLeuAlaGluPheMetGlyCysGlyPhe 272
Db      642 TTAAGGATGATACGAAACTCAGCTCAAAAGATTGGCTGAGTTTGTAGGTTGTCTCTTT 701
QY      273 ThrAspAspGluGluLysGlnGlyIleValAspClnIleValLysLeuCysSerPheAsp 292
Db      702 GTCGAGGAGGACATAGGAA-----GTTGAGGAGATAGTTTAAAGTGTGTAGCATTTGAT 755
QY      293 AsnLeuLysAsnGlnGlnValAsnLysAsn-----GlySerSer 305
Db      756 GTTTTGAAGAGTCATGAGGTGAACAAGCTCGAGGATGTCTCTGTGTTGGTTTCCATCTCCT 815
QY      306 TyrAsnSerLysIleAspAsnLysHisPhePheArgLysGlyGluValArgAspTrpAla 325
Db      816 TATAATTCA-----TTTTAGACAAGCAAGGTTGGTGACCATACA 857
QY      326 AsnTyrLeuThrSerGluMetIleLysLysLeuGluThrAlaGlyLys-IleAsnGluSe 345
Db      858 AATTACTTGATGATGAGTTTATCAAAAGAATTGATGACAGTTACTTAAGAGAAATTTTCC 917
QY      345 rGluLysHisLeuLeuSerAsn-----Ly 353
Db      918 AAATCTGGATTTGTTTATGGGATTTAGTGTCTTTAAATGAATAGTTTGTGAAGATTTAAG 977
QY      353 sAsnLeuThrAsnPheAsnLeu-----AsnAsnTyrCysGluLysSerAs 368
Db      978 ACCACTGTAAATTTCAAAATGTATGTTGCAATTTCAAGATTATATTGTTACTCGTACTT 1037
QY      368 nIleSerLeuCysLeuLeuSerTyrSerPheValIleAsnAsnPhelIleLeuLeuLysLy 388
Db      1038 AATATTGTATTGTGCTTTAAATCTCTCAATTCTCTCAATTCATGCTGCTGTTTTTCCAAAAAANA 1097
QY      388 sLysLysLys 391
Db      1098 AAAAAAANA 1107
RESULT 12
CC614744/c
LOCUS
DEFINITION
ACCESSION
VERSION
CC614744.1 GI:31976165
KEYWORDS
SOURCE
ORGANISM
Zea mays
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
Clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 877)
WhiteLaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.O.A., Rohlfing,T.,

```

Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.  
 Consortium for Maize Genomics  
 Unpublished (2002)  
 Other\_GSSs: OGUGQ80TH  
 Contact: Cathy Whitelaw  
 TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TP

Class: sheared ends.

FEATURES  
 Location/Qualifiers

1..877

/organism="Zea mays"

/mol\_type="genomic DNA"

/strain="B73"

/db\_xref="taxon:4577"

/clone="ZMMBMA0436M15"

/clone\_lib="ZM\_0.7-1.5\_KB"

/notes="vector: PBCSK-; Site 1: HincII; 0.7-1.5 kb  
 methylation filtered genomic DNA library"

ORIGIN

Alignment Scores:

Pred. No.: 3 628-39 Length: 877  
 Score: 451.00 Matches: 107  
 Percent Similarity: 54.73% Conservative: 55  
 Best Local Similarity: 36.15% Mismatches: 111  
 Query Match: 21.90% Indels: 23  
 DB: 29 Gaps: 8

US-09-854-122-16 (1-391) x CC614744 (1-877)

Qy 44 TyrArgGluLeuValSerSerLeuProSerSerAsnAspTyrTrpGlyAspThrMetArgLeu 63  
 Db 857 TACAGCATCTCGTATCCATCTGTCAGAGAGAGGGTGGTGCAGCGCTGGTACTG 798  
 Qy 64 TyrLysGlyPheTrpGlnMetGlyTyrLeuValProGly-----IleMetAla 79  
 Db 797 TACCAGAACTACTGG-----CTGACCCCGCGAGGCTCCAGCAGCATCATCCCC 750  
 Qy 80 PheGluAspAsnPhelYsAlaArgGluThrAspIleLeuThrThrLeuProLysAla 99  
 Db 749 GTGAAGAGAGCTCTACAGCCTCGCGCGAGCAGCATCTGCTCGTCACTACCGGAAATGC 690  
 Qy 100 GlyThrThrThrThrLysAlaLeuThrPheAlaIleLeuThrArgAspValAsnHisPro 119  
 Db 689 GGAACCACTGCTGAAGCTCTCGCTTCGCCATCAGACCCGC-----AGCCGCCAC 636  
 Qy 120 SerSerProThrHisProLeuLeuPhePheAsnProHisSerCysValGlnAsnLeuGlu 139  
 Db 635 GCTTTCGCGCGCCACCTCTGTCGCGCGCACCCTCAGGAGCTCGTGGCGCACCTTGAG 576  
 Qy 140 TyrLeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsn-GluSerPr 159  
 Db 575 CTGCGCGCCCTGGCGGGAGAC-----CTAGCGAGCTCGAAGAGCTAGTGCCTCCCTCCC 522  
 Qy 159 oArgLeuPheAlaGlyHisIleProTyrSerLeuLeuProAlaSerValLeuLysSerGl 179  
 Db 521 GAGGCTTCTCGGACTCACTACCTTCCCTTCTCCCGCGGACGTG-----TCGG 468  
 Qy 179 YThrLysIleLeuAsnIleSerArgAsnArgLysSerThrPheValSerPheThrLysPh 199  
 Db 467 CTACCGCTGTGTACTGTGTCGCGGACATCCCAAGAGCGCTCTGCTCGTGGTGGTCTT 408  
 Qy 199 eGlyAsnLeuIle-----AsnProAspLysLeuAspLeuGluLysSerVa 215  
 Db 407 CGCGAACCGCATGTGCGCAGAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 348  
 Qy 215 IAspIlePheAlaSerGlyIleSerPheCysGlyProGluTrpAsnPheGlnAlaGluPh 235  
 Db 347 GAGCATGTCTCGAGGGGTTCTCGCCCTTGGGCCCTTCTCTGGGAGCAGCTACCTCGAGTA 288

Qy 235 eThrAsnAlaAla---SerThrAsnSerAsnLeuLeuLeuSerTyrGluGluMetLe 254  
 Db 287 CTGGAGGCGGAGCTTGGAGAGCGCTCAGCATATCTCTGTTCTTGGGTACGAAGAGATCGT 228  
 Qy 254 uGluLysProValGluAsnValLysLeuAlaGluPheMetGlyCysGlyPheThrAs 274  
 Db 227 GCGAGACCCGCTGAAGGTGTCAAGACGCTTCCCGGCTTCTTCTGACCCGTTCACTGA 168  
 Qy 274 pAspGluGluLysGlnGlyIleValAspGluIleValLysLeuCysSerPheAsnLe 294  
 Db 167 GCGGGAAGAGAGAGGAGGAGTCCCGAGCAGGTGCGGAGGTGTCGAGTTCGAGGTGCT 108  
 Qy 294 uLysAsnGlnValAsnLysAsnGlySer-----SerTyrAsnSerLysIleAs 311  
 Db 107 CAGCGGCTCGAGAAACCAACACCGAGGTGTCGCTTTTGGAAACCATTTAGTCATTGG 48  
 Qy 311 pAsnLysHisPhePheArgLysGlyGluValArgAspTrpAlaAsn 326  
 Db 47 CAAGTCTACGTTCTTCAGGGAAGGCAAGGTGGGAGCTGGGAAAC 2

RESULT 13  
 CG196721

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES  
 source

1..807

/organism="Zea mays"

/mol\_type="genomic DNA"

/strain="B73"

/db\_xref="taxon:4577"

/clone="ZMMBMA0554D19"

/clone\_lib="ZM\_0.6-1.0\_KB"

/note="vector: PCR4-TOPO; Site\_1: EcoRI; 0.6-1.0 kb high  
 Cor selected genomic DNA library"

ORIGIN

Alignment Scores:

Pred. No.: 1 5e-37 Length: 807  
 Score: 436.00 Matches: 98  
 Percent Similarity: 53.82% Conservative: 43  
 Best Local Similarity: 37.40% Mismatches: 105  
 Query Match: 21.18% Indels: 16  
 DB: 29 Gaps: 7

US-09-854-122-16 (1-391) x CG196721 (1-807)

Qy 47 IleValSerSerLeuProSerSerAsnAspTyrTrpGlyAsp---ThrMetArgLeuTyrLys 65  
 Db 47 IleValSerSerLeuProSerSerAsnAspTyrTrpGlyAsp---ThrMetArgLeuTyrLys 65

```
Db 33 ATCATCCCTTCCTGCTCCCTCCGACACAGGTGGCCACCGTTCCTCTCCGACGGTACGCC 92
Qy 66 GlyPheTyrGlnMetGlyTyr---LeuValProGlyIleMetAlaPheGluAspAsnPhe 84
Db 93 AATTTTGGCTGCTGAGTAACTATTGAAGGAAGCGTCCCGGGCTTCATTCTCTGCTTC 152
Qy 85 LysAlaArgGluThrAspIlelelelelelelelelelelelelelelelelelelelele 104
Db 153 GAGCGAGAGACCCACCGAGCTGCTCTCGCAAGTTCCTCCCAAGTTCGGCACCACCTGGCTC 212
Qy 105 LysAlaLeuThrPheAlaIlelelelelelelelelelelelelelelelelelelelele 124
Db 213 AAGGCCCTGCTCCGCGAGCTGAAGCTTCCACGCTCCGCGCTTCGACGACGACCAT 272
Qy 125 ProLeuLeuPhePheAsnProHisSerCysValGlnAsnLeuGluTyrLeuTyrMetGly 144
Db 273 CCGCTGCGCCATTGCAACCTCATGCTGTGTCAGGTCTCTCGAA-----CTTGGT 323
Qy 145 ArgGluAsnThrMetProAspLeuAspMetLeuAsnGluSerProArgLeuPheAlaGly 164
Db 324 TTCACACGACAGAGATGATGCTCGAGGGCTC---CCGTCCCGGGTGTCTGGCCACG 380
Qy 165 HisIleProTyrSerLeuLeuProAlaSerValLeu-----LysSerGlyThr 180
Db 381 CACCTCCCTACTCCTGCTGCTCCCGGACATCACCGGGACGAGAGACCTCGGGGTGC 440
Qy 181 LysIlelelelelelelelelelelelelelelelelelelelelelelelelelelele 200
Db 441 CGCATGCTGTACGTGCGCGGAGCCCAAGGACATGCTGTCTCTACTGGCTGTTCAG 500
Qy 201 AsnLeuLeuAsnProAspLysLeuLeuAsp-----LeuGluLysSerVal 215
Db 501 AGGAAGGACGCGCCCTTGTGGGTGTGACCGCGCTGTTCCACGATCCGAGAGCCTTG 560
Qy 216 AspIlePheAlaSerGlyIleSerPheCysGlyProGluTyrAsnPheGlnAlaGluPhe 235
Db 561 GAGCTGTTCTGATGCGCTGCTGCGCGGCGCCACAGTGAACACGCTCTCCAGTAC 620
Qy 236 ---ThrAsnAlaAlaSerThrAsnSerAsnLeuLeuLeuLeuSerTyrGluLeuMetLeu 254
Db 621 TGGAGGAGAGCGTGGAGGCGCTGACAGGGTGTGTTCTCTGTTTACGAGAGGTGTG 680
Qy 255 GluLysProValGluAsnValLysLeuAlaGluPheMetGlyCysGlyPheThrAsp 274
Db 681 ATTGAGCCCGAGCGCGGCTGAGCAGATCGTGGAGTGTGATTCATGGCTGTGATCTCCCA 740
Qy 275 AspGluGluLysGlnGlyIleValAspGluIleValLysLeuLysSerPheAspAsnLeu 294
Db 741 GAAGAGAGAGACCCCGGGTGGTGAGCACCATCGTGGAGTGTGTAGCCTGGGCGAGCTG 800
Qy 295 LysAsn 296
Db 801 AGGGAC 806

RESULT 14
CG950535 855 bp DNA linear GSS 15-DEC-2003
LOCUS MBEG003TF mth2 Medicago truncatula genomic clone 51B5, genomic
DEFINITION survey sequence.
ACCESSION CG950535
VERSION CG950535.1 GI:39861114
KEYWORDS GSS.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE 1 (bases 1 to 855)
AUTHORS Town, C.D., Shetty, J., Koo, H. and Feldblum, T.F.
TITLE Sequencing of BAC ends from Medicago truncatula
JOURNAL Unpublished (2003)
COMMENT Other_GSSs: MBEG003TR
```

Contact: Chris Town  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
Seq primer: TGTAAACGACGCGCCAGT  
Class: BAC ends.

#### FEATURES

source  
1..855  
/organism="Medicago truncatula"  
/mol\_type="genomic DNA"  
/cultivar="genotype A17"  
/db\_xref="taxon:3880"  
/clone\_lib="51B5"  
/clone="51B5"  
/notes="Vector: pBeloBAC11; Site 1: HindIII; Site 2:  
HindIII; Cook, D.R. and Kim, D.J, unpublished"

#### ORIGIN

Alignment Scores:  
Pred. No.: 4,49e-37 Length: 855  
Score: 432.00 Matches: 101  
Percent Similarity: 58.49% Conservative: 54  
Best Local Similarity: 38.11% Mismatches: 102  
Query Match: 20.98% Indels: 9  
DB: 29 Gaps: 5  
US-09-854-122-16 (1-391) x CG950535 (1-855)  
Qy 36 GluAspSerLysMetTyrLysArgTyrArgGluIleValSerSerLeuProSerAsnAsp 55  
Db 66 CAGATGATGACTTATGACCAAGAAATGCAAGGATTTGATCCACCTTGCCTCAGAAAA 125  
Qy 56 TyrTyrGlyAspThrMetArgLeu---TyrLysGlyPheTyrGlnMetGlyTyrLeuVal 74  
Db 126 GCCTGGATTTTAAACCCACACCTATCAATATCAAGATTTTGGATTGCCAAAAAATTT 185  
Qy 75 ProGlyIleMetAlaPheGluAspAsnPheLysAlaArgGluThrAspIlelelelelele 94  
Db 186 CAAGGTGTTTATCTGTCAGAGACCTTTCAAGCCCTGATGATTCAGATATCATCTTAGTT 245  
Qy 95 ThrLeuProLysAlaGlyThrThrTyrThrLysAlaLeuThrPheAlaIlelelelelele 114  
Db 246 ACCGCTCCCAATACAGTACCCTTGGCTCAAGACCTTGTCTCATTGCTTAAACCCGC 305  
Qy 115 AspValAsnHisProSerSerProThrHisProleuLeuPhePheAsnProHisSerCys 134  
Db 306 AACAAATATCCAAATTATTCATAACACACCTCTTGTCTCATTTCCTCAATCCCATCTCT 365  
Qy 135 ValGlnAsnLeuGluTyrLeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMet 154  
Db 366 GTTCCCTTCTGGAGATTGACCTTACAGTGATAAGACTATGTCTCACTTAACCTTAAC 425  
Qy 155 LeuAsnGluSerProArgLeuPheAlaGlyHisIleProTyrSerLeuLeuProAlaSer 174  
Db 426 TTATCTCAT---CCAAGACTCTTCTCCACACATATTCCTTGTCTGTGTTGTCGCAAAATCT 482  
Qy 175 ValLeuLysSerGlyThrLysIlelelelelelelelelelelelelelelelelelele 194  
Db 483 GTGAAGATCACTCCAGGTTGTGTCTGTGTAGAGATCTCTTAAGACATGTTTGT 542  
Qy 195 SerPhe-----TyrLysPheGlyAsnLeuLeuAsnProAspLysLeu-----LeuAsp 210  
Db 543 TCACT-TTACTTTGGCATTTTCAAAACAAGTAAAGACACAGAGTAGAGGACCACTTCCA 601  
Qy 211 LeuGluLysSerValAspIlePheAlaSerGlyIleSerPheCysGlyProGluTyr--- 229  
Db 602 TTGAAGAGGCTTTTGAGAGTTTGTAGAGGAGTGTCTTATGAGCCCTTTGGAG 661  
Qy 230 AsnPheGlnAlaGluPheThrAsnAlaAlaSerThrAsnSerAsnLeuLeuLeuSer 249  
Db 662 CATGTATTAGTATTGGAGAAAAAATTTGAAAGGCCAGAGAGGTGATGTTTATAAAA 721

QY 250 TyrGluGluMetLeuGluLysProValGluAsnValLysLysLeuAlaGluPheMetGly 269  
Db 722 TTTGAAGAAATGAAATGAAACCTAGTTTTATTATTGAAAGAGATTGCTGAGTTCTGGGT 781  
QY 270 CysGlyPheThrAspAspGluGluGlyGlnGlyLeuValAspGluLysLeuCy 289  
Db 782 TGTCCATCTTTAATGAAGAAATGCAAGGGCATGATGATATATTAAATTTGTG 841  
QY 289 sSerPheAspAsn 293  
Db 842 TAGTTTGGAGAC 854

RESULT 15  
CG269067  
LOCUS CGXCR72TV ZM\_0.7\_1.5\_KB Zea mays genomic clone ZMMBMA0650K24,  
DEFINITION genomic survey sequence.  
ACCESSION CG269067  
VERSION CG269067.1 GI:34181208  
KEYWORDS GSS.  
SOURCE Zea mays  
ORGANISM Zea mays

REFERENCE 1 (bases 1 to 966)  
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,  
Resnick,A., Frazer,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,  
Citek,R.W., Nurnberg,A., Robbins,D. and Lakey,N.  
TITLE Consortium for Maize Genomics  
JOURNAL Unpublished (2002)  
COMMENT Other GSSs: OGXCR72TH  
Contact: Cathy Whitelaw  
TIGR Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whitelaw@tigr.org  
Seq primer: TF  
Class: sheared ends.

FEATURES  
source  
1..966  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBMA0650K24"  
/clone\_lib="ZM\_0.7\_1.5\_KB"  
/notes="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb  
methylation filtered genomic DNA library"

ORIGIN  
Alignment Scores:  
Pred. No.: 2,14e-36 Length: 966  
Score: 426.50 Matches: 108  
Percent Similarity: 53.57% Conservative: 57  
Best Local Similarity: 35.06% Mismatches: 117  
Query Match: 20.71% Indels: 26  
DB: 29 Gaps: 10

US-09-854-122-16 (1-391) x CG269067 (1-966)  
QY 14 LeuileMetAlaGlylleLeuAlaLeuGluLys-----CysPhe-GlySe 28  
Db 25 TTAATCCACCGGCCATCGTTCTTCTGAACAAAGCCATGGCGCAAGCTGCTCCGGCTC 84  
QY 28 rLysAsnGluGlnGluLys-----GluGluAspSerLysMerTyrIysArgTy 44  
Db 85 CAGCAAGATCGACACCAAGCTCCACCTCAGCGGACGAGGCCGCTCAACCAAGAGCTA 144  
QY 44 rArgGluileValSerSerLeuProSerAsnAspTyrTrpGlyAspThrMetArgLeuTy 64

Db 145 CGGCGACCTGGTATCCACTCTCCCGGTAGAGCGGGTGGGTATCGCTCGTGCGAG---TA 201  
QY 64 rLysGlyPheThrGlnMetGlyTyrLeuValProGly-----lleMetAlaPh 80  
Db 202 CCGAAGAACTACTGG-----CTGAACCCCGGAGGCTCCAGAACATCATCCCCGT 249  
QY 80 eGluAspAsnPhelYsAlaArgGluThrAspIlelleLeuThrThrLeuProLysAlaGl 100  
Db 250 GAAAGAGTCTTACAAGCTCGCGCGACGACATCTCTCGCCACTACCTGAAATGCGG 309  
QY 100 yThrThrTrpThrLysAlaLeuThrPheAlaIleLeuThrArgaspValAsnHisPro-- 119  
Db 310 GACCACGTGGTGAAGGCGCTGGCCCTTCCCATCAGACCCCGCGCGCCACCAACAGGC 369  
QY 120 -----SerSerProThrHisProLeuLeuPhePheAsnProHisSerCysValGlnAsnLe 138  
Db 370 CTTCCGCGCGCGCGACCACTCCCTCTCAGCTGACCCGCGAGGAGGCGGTGCCCACT 429  
QY 138 uGluTyrLeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsnGluSe 158  
Db 430 GGAGGTGCACACGCCAGGGCGAG-----GGCTGGCAGACATCGAGAGGCTG---CCGTC 480  
QY 158 rProArgLeuPheAlaGlyHisIleProTyrSerLeuLeuProAlaSerValLeuLysSe 178  
Db 481 GCCGAGGTGCTCGGAGCGACCTGCCCTCTCCTCTGCTCCGCGCGCGGTGGCCACCTC 540  
QY 178 rGlyThrLysIlelleAsnIleSerArgAsnArgLysSerThrPheValSerPheTrpLy 198  
Db 541 CGGCTCGCGCGTGTGTACTCTGCGCGCAGCCCAAGGACGTGTCTGCTCTGCTGCGCA 600  
QY 198 sPheGlyAsnLeuIleAsnProAspLysLeu---LeuAspLeuGluLysSerValAspIl 217  
Db 601 CTTCTGTAAGGCGATGCGCGGGGCGAGTCCCGGTGGAGTGGACGCGCGGTGGCCAT 660  
QY 217 ePheAlaSerGlylleSerPheCysGlyProGluTrpAsnPheGlnAlaGluPhe---Th 236  
Db 661 GTTCTGCGAGGGGTCTCGCCTTTCCGGCGCGCTCTGGAGCACTACTCTGAGTACTGAA 720  
QY 236 rAsnAlaAlaSerThrAsnSerAsnLeuLeuLeuSerTyrGluGluMetLeuGluLy 256  
Db 721 GGAGAGCTGGCGAGGCGCTGAGCGGGTCTTTCTCTAGGTACGAGGAGATGGTGGCGGA 780  
QY 256 sProValGluAsnValLysLysLeuAlaGluPheMetGlyCysGlyPheThrAspAspGl 276  
Db 781 CCGGTGCGTGGTGGAGGCGCTCGCGCGGTCTTCTCGCGCGTCCGCTTCCCGACGAGGA 840  
QY 276 uGluLysGlnGlylleValAspGluIleValLysLeuCysSerPheAspAsnLeuLysAs 296  
Db 841 AGAGGGCGAGGGGTCCCGGAGGAGATTGTGAGGCTGTGCAGCTTCGAGATGCTGAGCGG 900  
QY 296 nGlnGlnValAsnLysAsnGly 303  
Db 901 CTTGGAAAAAGCATCGGACCGGA 922

Search completed: March 20, 2004, 06:21:51  
Job time : 3322 secs